Validierung und Weiterentwicklung neuer Methoden zur Bestimmung der Faktorenzahl in explorativen Faktorenanalysen

Inaugural-Dissertation

zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultät
der Heinrich-Heine-Universität Düsseldorf

vorgelegt von

Nils Brandenburg
aus Duisburg

Düsseldorf, September 2022

aus dem Institut für Experimentelle Psychologie
der Heinrich-Heine-Universität Düsseldorf

Gedruckt mit der Genehmigung der
Mathematisch-Naturwissenschaftlichen Fakultät der
Heinrich-Heine-Universität Düsseldorf

Berichterstatter:

1. Prof. Dr. Axel Buchner
2. Prof. Dr. Jochen Musch

Tag der mündlichen Prüfung: 16.12.2022

Inhaltsverzeichnis

Zusammenfassung

Im Kontext explorativer Faktorenanalysen existieren zahlreiche Methoden zur Bestimmung der Faktorenzahl. Simulierte Zufallsdatensätze unter Faktorenmodellen mit bekannter Faktorenzahl eignen sich zur Validierung solcher Methoden, indem die Genauigkeit ermittelt wird, mit der Methoden die bekannte Faktorenzahl reproduzieren. In der vorliegenden Dissertation werden die Methoden *Next Eigenvalue Sufficiency Test* (NEST; Achim, 2017) und *Exploratory Graph Analysis* (EGA; Golino & Epskamp, 2017; Golino et al., 2020) behandelt. Beide Methoden wurden zuvor mit dem Befund validiert, die Genauigkeit der *Parallelanalyse* (PA; Horn, 1965) als Goldstandardreferenz in Simulationen übertroffen zu haben. Für die vorliegende Dissertation wurden NEST und EGA erstmals in Simulationen miteinander verglichen. Die Ergebnisse zeigen, dass NEST, EGA und auch PA allesamt vergleichbar genau waren, wenn simulierte Variablen jeweils nur einen Faktor anzeigten (Simulation 1a). Darüber hinaus wurden NEST und EGA erstmals in Simulationsbedingungen untersucht, in denen simulierte Variablen jeweils mehrere Faktoren anzeigten. NEST und PA waren weiterhin vergleichbar genau, wenn simulierte Variablen jeweils zwei Faktoren anzeigten; EGA war hingegen deutlich ungenauer als NEST und PA (Simulation 1b). Das Genauigkeitsdefizit von EGA wurde in weiteren Simulationen anhand von Circumplexmodellen bestätigt, in denen Variablen ebenfalls jeweils mehrere Faktoren anzeigten (Simulationen 2a bis 2f). Die Ergebnisse zeigen insgesamt, dass NEST weniger restriktive Annahmen zur faktoriellen Struktur gemessener Variablen voraussetzt als EGA und daher für explorative Faktorenanalysen besser geeignet ist. Außerdem wird in der vorliegenden Dissertation ein Vergleich verschiedener Implementierungen von NEST für Anwendungen mit kategorialen Variablen behandelt. Dazu wurde die ursprüngliche Implementierung basierend auf Pearson-Korrelationen mit einer modifizierten Implementierung basierend auf polychorischen Korrelationen verglichen (Simulation 3). NEST mit polychorischen Korrelationen war die genaueste Methode für binäre Variablen, setzte jedoch große Stichproben ($N = 500$) voraus. NEST mit Pearson-Korrelationen war die genaueste Methode für Variablen mit vier Antwortkategorien, setzte jedoch homogene Item-Schwierigkeiten voraus. Insgesamt wird die Validierung von NEST durch die vorliegende Dissertation um zusätzliche Methoden- und Implementierungsvergleiche erweitert. Die Ergebnisse grenzen ein, unter welchen Bedingungen NEST Vorteile gegenüber PA zeigt und unter welchen nicht.

Abstract

In the context of exploratory factor analysis, there are several methods to determine the number of factors. Random data sets simulated under factors models with known numbers of factors may serve the validation of such methods in that the accuracy with which methods recover the known number of factors is investigated. The present work concerns the methods *Next Eigenvalue Sufficiency Test* (NEST; Achim, 2017) and *Exploratory Graph Analysis* (EGA; Golino & Epskamp, 2017; Golino et al., 2020). Both methods have been reported to surpass the accuracy of *Parallel Analysis* (PA; Horn, 1965)—the common gold standard benchmark—in simulations. The present work provides a first direct comparison of NEST and EGA in a series of simulations. The results show that NEST, EGA, and also PA were all similar in accuracy when simulated variables indicated one factor each (Simulation 1a). Furthermore, NEST and EGA were tested in conditions where simulated variables indicated multiple factors each. NEST and PA continued to be similar in accuracy when variables indicated two factors each whereas EGA was considerably less accurate than NEST and PA (Simulation 1b). The deficit in accuracy of EGA continued to show in additional simulations targeting circumplex models in which variables again indicated multiple factors each (Simulations 2a to 2f). In summary, the present work reveals that NEST is less restrictive than EGA in its assumptions concerning the factorial structure of analyzed data and hence better suited for application in exploratory factor analysis. In addition, the present work includes a comparison of different implementations of NEST with respect to analyses of categorical variables. The original implementation of NEST based on Pearson correlations was compared to a modified implementation based on polychoric correlations (Simulation 3). NEST with polychoric correlations was the most accurate method for binary variables, albeit limited to large samples ($N = 500$). NEST with Pearson correlations was the most accurate method for variables with four response categories, albeit limited to homogeneous item difficulties. All in all, the present work extends the previous validation of NEST through comparison to an extended set of competing methods and comparisons of different implementations of NEST. The present results narrow down the conditions in which NEST shows superior accuracy to PA and the conditions in which NEST does not.

Einleitung

Die Berechnung der Korrelationen gemessener Variablen ist ein elementares Vorgehen in der empirischen Psychologie. Daran anknüpfend können weiterführende Forschungsfragen die Ursache beobachteter Korrelationen behandeln. Der Begriff *Faktorenanalyse* bezeichnet eine Familie statistischer Methoden, in denen beobachtete Korrelationen als Funktion latenter Merkmale modelliert werden (siehe Cattell, 1978; Jöreskog, 1978; Thurstone, 1931; Widaman, 2018). Die latenten Merkmale zur Erklärung beobachteter Korrelationen werden dabei als *gemeinsame Faktoren* der gemessenen Variablen bezeichnet. Dem Ansatz der Faktorenanalyse zufolge korrelieren gemessene Variablen, wenn sie dieselben gemeinsamen Faktoren operationalisieren.

Faktorenanalysen sind in diversen psychologischen Forschungsbereichen verbreitet. Beispielsweise spielen Faktorenanalysen eine zentrale Rolle bei der Entwicklung psychologischer Tests (Henson & Roberts, 2006; McNeish & Wolf, 2020; Schmitt, 2011; Ziegler & Hageman, 2015). In der Testentwicklung ermöglichen Faktorenanalysen, Items auf gemeinsame Faktoren abzubilden. Aus einer solche Abbildung ergibt sich, welche Items welches latente Merkmal messen und wie sich Items dadurch in Subskalen gliedern lassen (Beispiele: Jonason & Webster, 2010; McCrae et al., 1996; Steer et al., 1999; Watson et al., 1988). Darüber hinaus sind Faktorenanalysen auch in der Grundlagenforschung in der Psychopathologie relevant (Hallquist et al., 2021; Reise, 2012). Die Identifikation gemeinsamer Faktoren von Beschwerden trägt dazu bei, das komorbide Auftreten von Störungsbildern zu erklären (Eaton et al., 2013; Forbush & Watson, 2013; Krueger & Markon, 2006; Watson, 2005) und Störungsbilder in abgrenzbare Gruppen einzuordnen (Kotov et al., 2011; Sharp et al., 2015).

In der vorliegenden Dissertation werden Fragestellungen zur optimalen Durchführung von Faktorenanalysen aus einer methodenwissenschaftlichen Perspektive betrachtet. Der umfangreichen Historie zum Trotz beinhalten Faktorenanalysen je nach Anwendungsform fundamentale Probleme, deren Lösungen bis heute umstritten sind. In den folgenden Abschnitten werden die mathematischen Grundlagen der Faktorenanalyse dargelegt. Aus den Grundlagen ergeben sich die konkreten Probleme, die in der vorliegenden Dissertation behandelt werden.

Das univariate Faktorenmodell

Faktorenanalysen werden primär in Kontexten mehrerer gemessener Variablen angewendet. Das zugrundeliegende lineare Modell – im Nachfolgenden *Faktorenmodell* genannt – hat jedoch auch eine univariate Form, die die Kernaspekte des Faktorenmodells veranschaulicht und aus der sich die multivariate Form ergibt. Die grundlegenden Annahmen für Faktorenmodelle sind in der vorliegenden Dissertation, dass Zusammenhänge zwischen gemessenen Variablen und gemeinsamen Faktoren linear sind, und dass sämtliche Variablenrealisierungen unabhängig und identisch verteilt sind. Das univariate lineare Faktorenmodell lässt sich unter diesen Annahmen folgendermaßen ausdrücken:

$$z_{ij} = \sum_{k=1}^{m} \left(\lambda_{jk}\eta_{ik}\right) + \zeta_{ij} \tag{1}$$

In Gleichung 1 bezeichnen z_{ij} den Wert der gemessenen Variablen z_j für die Beobachtung i, η_{ik} den Wert des gemeinsamen Faktors η_k für die Beobachtung i, λ_{jk} den Regressionskoeffizienten der Variablen z_j für den Faktor η_k, m die Zahl der gemeinsamen Faktoren und ζ_{ij} den Wert eines variablenspezifischen Anteils ζ_j der Variablen z_j für die Beobachtung i. Die Modellparameter λ_{jk} werden im Rahmen der Faktorenanalyse geschätzt und als *Faktorladungen* bezeichnet. Gleichung 1 gibt an, dass im Faktorenmodell zwischen dem Term ζ_{ij} zur Beschreibung variablenspezifischer Varianzanteile und einem komplementären Term $\sum_{k=1}^{m}\left(\lambda_{jk}\eta_{ik}\right)$ zur Beschreibung geteilter Varianzanteile unterschieden wird.

Der Anspruch des Faktorenmodells, geteilte und ungeteilte Varianz zu trennen und nur die geteilte Varianz durch gemeinsame Faktoren zu modellieren, wird auch in folgender Varianzzerlegung deutlich:

$$\mathrm{var}\left(z_j\right) = \mathrm{var}\left(\sum_{k=1}^{m}\left(\lambda_{jk}\eta_k\right)\right) + \mathrm{var}\left(\zeta_j\right) \tag{2}$$

Der Term $\mathrm{var}\left(\sum_{k=1}^{m}\left(\lambda_{jk}\eta_k\right)\right)$ der Gleichung 2 beschreibt die geteilte Varianz der Variablen z_j. Die geteilte Varianz wird in Faktorenanalysen üblicherweise als *Kommunalität* bezeichnet wird. Mit der Zusatzannahme, dass alle gemeinsamen Faktoren unabhängig voneinander sind, kann Gleichung 2 äquivalent angegeben werden durch:

$$\mathrm{var}(z_j) = \sum_{k=1}^{m} \left(\lambda_{jk}^2 \mathrm{var}(\eta_k)\right) + \mathrm{var}(\zeta_j) \quad (3)$$

Das multivariate Faktorenmodell

Die multivariate Extension der Gleichung 3 beschreibt nicht die Varianz einer einzelnen gemessenen Variablen, sondern bei p gemessenen Variablen die $p \times p$-Varianz-Kovarianzmatrix Σ und deren Zerlegung in geteilte und ungeteilte Varianzanteile. Für Σ als multivariate Extension von $\mathrm{var}(z_j)$ ergibt sich folgende Extension der Gleichung 3:

$$\Sigma = \Lambda\Psi\Lambda^{\mathrm{T}} + \Theta \quad (4)$$

In Gleichung 4 bezeichnen Λ die $p \times m$-Faktorladungsmatrix, Ψ die $m \times m$-Varianz-Kovarianzmatrix der gemeinsamen Faktoren und Θ die $p \times p$-Varianz-Kovarianzmatrix der ungeteilten Varianzanteile. Die Zusatzannahme, dass gemeinsame Faktoren unabhängig voneinander sind, ist in Faktorenanalysen nicht zwingend erforderlich, sodass Ψ nicht notwendigerweise eine Diagonalmatrix ist. Im weiteren Verlauf wird – wie in Faktorenanalysen üblich – angenommen, dass die ungeteilten Varianzanteile aller Variablen unabhängig voneinander und unabhängig von gemeinsamen Faktoren sind. Außerdem wird angenommen, dass alle gemessenen Variablen und alle gemeinsamen Faktoren standardisiert sind und somit eine Varianz von 1.00 haben; dadurch stellen Σ und Ψ Korrelationsmatrizen dar[1].

Die Faktorladungsmatrix Λ gibt an, welche gemessenen Variablen mit welchen gemeinsamen Faktoren assoziiert sind. Die Berechnung von Λ ist daher der zentrale Schritt in Faktorenanalysen zur Zuordnung gemessener Variablen zu gemeinsamen Faktoren. Die inhaltliche Deutung eines gemeinsamen Faktors ergibt sich aus den gemessenen Variablen, die den Faktor der Matrix Λ zufolge anzeigen (Fabrigar et al., 1999; Jöreskog, 1978; Widaman, 2018).

[1] Wenn Σ und Ψ Korrelationsmatrizen sind und ungeteilte Varianzanteile unabhängig sind, folgt für Gleichung 4, dass $\Theta = \mathrm{I}_{p\times p} - \mathrm{diag}(\Lambda\Psi\Lambda^{\mathrm{T}})$.

Explorative Faktorenanalysen und das Faktorenzahlproblem

In der vorliegenden Dissertation wird ein Sonderfall der Faktorenanalyse behandelt, bei dem ein Faktorenmodell für gemessene Variablen ohne Annahmen zur Zuordnung gemessener Variablen zu gemeinsamen Faktoren geschätzt wird. Ein Faktorenmodell ohne Zuordnungsannahmen zu schätzen ist insbesondere in frühen Stadien der Analyse gemessener Variablen sinnvoll – beispielsweise mit einer ersten Stichprobe eines Testentwurfes –, wenn Hypothesen zu gemeinsamen Faktoren zunächst generiert werden sollen (Auerswald & Moshagen, 2019; Clark & Bowles, 2018; Hallquist et al., 2021; Widaman, 2018; Ziegler & Hagemann, 2015). Dieser Sonderfall der Faktorenanalyse wird daher als *explorative Faktorenanalyse* bezeichnet.

In explorativen Faktorenanalysen bestehen distinkte Probleme bei der Schätzung eines Faktorenmodells (siehe Cattell, 1978; Fabrigar et al., 1999). Das spezielle Problem, das Gegenstad der vorliegenden Dissertation ist, betrifft die Bestimmung der Anzahl gemeinsamer Faktoren m für ein Faktorenmodell (siehe oben). Dieses Problem wird als *Faktorenzahlproblem*[2] bezeichnet. Das übergeordnete Ziel in der vorliegenden Dissertation ist es, zur Lösung des Faktorenzahlproblems in explorativen Faktorenanalysen beizutragen.

Das Faktorenzahlproblem tritt auf, wenn bestehende Erkenntnisse zur faktoriellen Struktur gemessener Variablen nicht ausreichen, um Annahmen zur Anzahl bedeutsamer Faktoren zu rechtfertigen. In der Literatur zum Faktorenzahlproblem besteht der Konsens, dass die Bestimmung einer geeigneten Faktorenzahl die Ergebnisqualität explorativer Faktorenanalysen maßgeblich beeinflusst (siehe Auerswald & Moshagen, 2019; Henson & Roberts, 2006; Li et al., 2020; Preacher et al., 2013; Schmitt, 2011; Zwick & Velicer, 1986). Die Bestimmung der Faktorenzahl betrifft inhaltliche und statistische Anforderungen an Faktorenmodelle. Eine Anforderung ist, vor der Schätzung des Faktorenmodells abzusichern, dass Modellparameter geschätzt werden für alle theoretisch bedeutsamen Faktoren, die bei der Erklärung beobachteter Korrelationen einer Rolle spielen (Cosemans et al., 2022; Ruscio & Roche, 2012; Zwick & Velicer, 1986). Eine weitere Anforderung ist,

[2] In Faktorenanalysen wird zwischen »gemeinsamen Faktoren« (für geteilte Varianz) und »variablenspezifischen Faktoren« (für ungeteilte Varianz) unterschieden. Das Faktorenzahlproblem bezieht sich exklusiv auf die Bestimmung der Anzahl gemeinsamer Faktoren. Da die vorliegende Dissertation das Faktorenzahlproblem behandelt und sich somit primär auf gemeinsame Faktoren bezieht, wird im weiteren Verlauf vereinfachend der Begriff »Faktor« für gemeinsame Faktoren verwendet.

das Faktorenmodell auf die theoretisch bedeutsamen Faktoren zu begrenzen und keine weiteren Faktoren darüber hinaus zu berücksichtigen (Auerswald & Moshagen, 2019; Braeken & van Assen, 2017; Fava & Velicer, 1992; Lorenzo-Seva et al., 2011). Die optimale Faktorenzahl (vgl. Fabrigar et al., 1999; Preacher et al., 2013) ist demzufolge die Faktorenzahl, durch die diese Anforderung balanciert sind (Achim, 2017; Revelle & Rocklin, 1979).

Im Gegensatz zur Bedeutung des Faktorenzahlproblems für die Qualität von Faktorenmodellen besteht kein Konsens zur Lösung des Faktorenzahlproblems. Die Literatur zum Faktorenzahlproblem enthält zahlreiche Methoden, die zur Bestimmung der optimalen Faktorenzahl entwickelt wurden. Die Entwicklung von Methoden zur Bestimmung der Faktorenzahl ist ein historisches, aber weiterhin aktives Forschungsfeld (siehe Achim, 2017; Braeken & van Assen, 2017; Cattell, 1966; Golino & Epskamp, 2017; Goretzko & Bühner, 2020; Green et al., 2012; Guttman, 1954a; Horn, 1965; Lorenzo-Seva et al., 2011; Ruscio & Roche, 2012; Velicer, 1976). Die Forschungsfragen der vorliegenden Dissertation basieren auf aktuellen Entwicklungen zum Faktorenzahlproblem.

Die erste beigefügte Einzelarbeit (Brandenburg & Papenberg, 2022) ist eine vergleichende Arbeit zu den veröffentlichten Methoden von Achim (2017) und Golino und Epskamp (2017). In der ersten Einzelarbeit werden Lösungen[3] der beiden Methoden für das Faktorenzahlproblem verglichen und die Vereinbarkeit der Methoden mit den grundlegenden Annahmen explorativer Faktorenanalysen thematisiert. Die zweite beigefügte Einzelarbeit (Brandenburg, 2022) ist eine weiterführende Arbeit zur Methode von Achim (2017), in der die Lösungen der Methode für kategoriale Variablen behandelt werden. In der zweiten Einzelarbeit wird die ursprüngliche Implementierung der Methode von Achim verglichen mit einer modifizierten Implementierung, die speziell zur Bestimmung der Faktorenzahl für kategoriale Variablen entwickelt wurde. In den nachfolgenden Zusammenfassungen der Einzelarbeiten[4] werden die untersuchten Methoden zur Bestimmung der Faktorenzahl eingeführt. Dabei werden auch die konkreten Problemstellungen definiert, anhand derer die Methoden für die vorliegende Dissertation untersucht wurden.

[3] Als »Lösung« wird in der vorliegenden Dissertation die Faktorenzahl bezeichnet, die eine Methode zur Bestimmung der Faktorenzahl für einen gegebenen Datensatz bestimmt.
[4] Die Referenzen »Brandenburg & Papenberg (2022)« und »Brandenburg (2022)« in den nachfolgenden Abschnitten beziehen sich auf die erste bzw. zweite beigefügte Einzelarbeit.

Simulationsstudien im Kontext des Faktorenzahlproblems

Zur Validierung von Methoden zur Bestimmung der Faktorenzahl werden üblicherweise Simulationsstudien durchgeführt (Beispiele: Auerswald & Moshagen, 2019; Cosemans et al., 2022; Lim & Jahng, 2019). Das grundlegende Simulationsparadigma wurde für die vorliegende Dissertation übernommen und wird daher in den folgenden Abschnitten erläutert. Simulationsstudien im Kontext des Faktorenzahlproblems werden nach dem Monte-Carlo-Prinzip durchgeführt, um die Lösungen von Methoden zur Bestimmung der Faktorenzahl anhand computergenerierter Zufallsdatensätze zu untersuchen. Simulierte Datensätze bieten gegenüber empirischen Datensätzen den entscheidenden Vorteil, dass der datengenerierende Prozess in Simulationen vollumfänglich und eindeutig bekannt ist.

In Simulationsstudien werden Faktorenmodelle gemäß Gleichung 4 spezifiziert, die eine bestimmte Faktorenzahl enthalten und jeweils eine Populationskorrelationsmatrix implizieren. Aus den modellimplizierten Populationen kann dann eine beliebige Anzahl von zufallsgenerierten Stichproben gezogen werden. Untersuchte Methoden zur Bestimmung der Faktorenzahl können dann auf die simulierten Datensätze angewendet werden. Das Validierungskriterium für Methoden ist die Reproduktion der Faktorenzahl der Faktorenmodelle, die die Populationen der simulierten Datensätze implizieren und deren Faktorenzahl daher als optimal betrachtet werden kann. Da die Lösungen der Methoden die optimale Faktorenzahl sowohl unter- als auch überschätzen[5] können, bieten Simulationsstudien darüber hinaus die Möglichkeit, Methoden hinsichtlich der Häufigkeiten von Unter- und Überschätzungen der optimalen Faktorenzahl zu untersuchen.

Die Merkmale der Faktorenmodelle und der darunter simulierten Datensätze wurden in den Simulationen der vorliegenden Dissertation jeweils nach zwei Gesichtspunkten spezifiziert. Zum einen wurden Abhängigkeiten der Lösungen untersuchter Methoden durch gezielte Manipulationen ausgewählter Datenmerkmale untersucht. Zum anderen wurden diverse Datenmerkmale

[5] In der vorliegenden Dissertation werden Unter- und Überschreitungen der optimalen Faktorenzahl als Unter- bzw. Überschätzungen bezeichnet, obwohl es sich dabei strenggenommen nicht um Schätzungen im Sinne einer Parameterschätzung handelt.

manipuliert, um die Methoden in heterogenen simulierten Anwendungsfällen zu untersuchen.

Methodenvergleich

In der jüngeren Entwicklung neuer Methoden zur Bestimmung der Faktorenzahl wurden neue Methoden in Simulationsstudien untersucht und dabei mit Referenzmethoden verglichen (siehe Achim, 2017; Braeken & van Assen, 2017; Golino et al., 2020; Goretzko & Bühner, 2020; Green et al., 2012; Lorenzo-Seva et al., 2011; Ruscio & Roche, 2012). Die prominenteste Referenzmethode für neue Methoden ist die *Parallelanalyse* (PA; Horn, 1965), die sich in der Vergangenheit als Goldstandardreferenz etabliert hat (Auerswald & Moshagen, 2019; Garrido et al., 2013; Goretzko et al., 2021; Lubbe, 2019). Der Status von PA als Goldstandard beim Faktorenzahlproblem basiert auf älteren Simulationsstudien, in denen PA die optimale Faktorenzahl häufiger als andere Methoden reproduzierte (Fabrigar et al., 1999; Velicer et al., 2000; Zwick & Velicer, 1986).

Achim (2017) veröffentlichte den *Next Eigenvalue Sufficiency Test* (NEST) und berichtete, dass NEST in einer Simulationsstudie die optimale Faktorenzahl häufiger reproduzierte als PA. Golino und Epskamp (2017) veröffentlichten zur gleichen Zeit die *Exploratory Graph Analysis* (EGA), die ebenfalls in einer Simulationsstudie die optimale Faktorenzahl häufiger reproduzierte als PA (Golino et al., 2020). Die beiden Simulationsstudien trugen damit zum Umgang mit dem Faktorenzahlproblem bei, dass neue Methoden zur Lösung entwickelt wurden und ihre Anwendbarkeit mit simulationsbasierten Validierungen untermauert wurde. Diese Beiträge sind jedoch dahingehend limitiert, dass NEST und EGA bisher nicht miteinander verglichen wurden. Diese Limitation wird in der vorliegenden Dissertation in Form eines simulationsbasierten Vergleiches in den Simulationen 1a bis 2f adressiert.

Kreuzladungen

Die bisherigen Simulationsstudien zu NEST (Achim, 2017) und EGA (Golino et al., 2020) wurden so gestaltet, dass simulierte Variablen nur eine starke Faktorladung auf jeweils einen Faktor zeigten – auch dann, wenn die spezifizieren Faktorenmodelle mehrere Faktoren enthielten. Faktorenmodelle, in denen gemessene Variablen jeweils einen Faktor anzeigen, bilden jedoch einen Sonderfall (Acton &

Revelle, 2004; Jöreskog, 1966; Revelle & Rocklin 1979), da Faktorenmodelle theoretisch mehrere Faktorladungen pro gemessene Variable erlauben. Gemessene Variablen mit substantiellen Faktorladungen auf mehrere Faktoren zeigen folglich mehr als einen Faktor an. Mehrfache Faktorladungen gemessener Variablen werden üblicherweise als *Kreuzladungen* bezeichnet.

Kreuzladungen eignen sich besonders gut zur Validierung von Methoden im Kontext explorativer Faktorenanalysen: Der Verzicht auf Annahmen zur Zuordnung gemessener Variablen zu Faktoren in explorativen Faktorenanalysen impliziert, dass keine inhärenten Annahmen zu Kreuzladungen bestehen (Achim, 2020). Folglich würde es die Anwendbarkeit von Methoden zur Bestimmung der Faktorenzahl in explorativen Faktorenanalysen limitieren, wenn sie die optimale Faktorenzahl nur dann zuverlässig anzeigen können, wenn gemessene Variablen jeweils nur einen Faktor anzeigen. Aus diesem Grund wurde für die vorliegende Dissertation erstmals untersucht, wie robust NEST und EGA gegen substantielle Kreuzladungen sind, durch die gemessene Variablen mehreren Faktoren zuzuordnen sind. Dazu wurden Faktorenmodelle mit Kreuzladungen spezifiziert, um Datensätze unter ihnen zu simulieren. Auf diese Weise sollte untersucht werden, in welchem Umfang NEST und EGA durch Kreuzladungen limitiert sind und ob sich daraus Methodenpräferenzen für explorative Faktorenanalysen ergeben.

Next Eigenvalue Sufficiency Test

Im Feld der Methoden zur Bestimmung der Faktorenzahl lässt sich NEST in eine Klasse von Methoden einordnen, die auf einer Eigenwertzerlegung der Stichprobenkorrelationsmatrix basieren (Auerswald & Moshagen, 2019; Golino & Epskamp, 2017; Li et al., 2020; Lorenzo-Seva et al., 2011; Revelle & Rocklin, 1979; Ruscio & Roche, 2012; Schmitt, 2011). Die Betrachtung von Eigenwerten der Stichprobenkorrelationsmatrix beim Faktorenzahlproblem begründet sich darin, dass der Betrag der Eigenwerte interpretiert werden kann als Indikator, um welchen Betrag die Summe der Kommunalität aller Variablen durch die Hinzunahme eines weiteren Faktors inkrementiert wird (Auerswald & Moshagen, 2019; Braeken & van Assen, 2017). Vereinfacht formuliert besteht bei einem Faktorenmodell eine Korrespondenz zwischen den m Faktoren des Modells und den m größten Eigenwerten einer Stichprobenkorrelationsmatrix unter dem Modell.

NEST nutzt die Eigenwerte der Stichprobenkorrelationsmatrix eines Datensatzes für eine sequentielle Testung der Nullhypothese, dass k Faktoren für ein

passendes Faktorenmodell ausreichen. Der Laufindex k für die Faktorenzahl beginnt bei 0 und wird mit jeder Zurückweisung der Nullhypothese um 1 inkrementiert. Die Teststatistik beim Test des Faktorenmodells mit k Faktoren ist der Eigenwert der Stichprobenkorrelationsmatrix an der Stelle $k + 1$ (Eigenwerte sind dabei der Größe nach absteigend geordnet). Für jeden Test wird ein Faktorenmodell mit k Faktoren für den gegebenen Datensatz berechnet. Unter dem entsprechenden Faktorenmodell werden zufallsgenerierte Referenzdatensätze simuliert. Im Fall $k = 0$ werden Referenzdatensätze aus einer Population unabhängiger Variablen gezogen. Für jeden Referenzdatensatz wird ebenfalls eine Stichprobenkorrelationsmatrix berechnet. Das getestete Faktorenmodell mit k Faktoren ist der wahre datengenerierende Prozess der Referenzdatensätze. Daher bilden die Eigenwerte der Referenzdatensätze an der Stelle $k + 1$ die Stichprobenverteilung des getesteten Eigenwertes unter der Nullhypothese ab, dass k Faktoren ausreichen, um Referenzdatensätze zu simulieren, die sich nicht vom analysierten Datensatz unterscheiden. Wenn der getestete Eigenwert die simulierte Stichprobenverteilung signifikant übersteigt, wird die Nullhypothese, dass k Faktoren ausreichen, zurückgewiesen (siehe Brandenburg & Papenberg, 2022, S. 3, für eine technische Einführung in NEST inklusive einer Zusammenfassung des non-parametrischen Hypothesentests in NEST).

Die Robustheit von NEST gegen substantielle Kreuzladungen erfordert gezielte Untersuchungen, da Kreuzladungen getestete Eigenwerte reduzieren können (siehe Brandenburg & Papenberg, 2022, S. 3, für eine Erläuterung dieses Effektes basierend auf Implikationen der Gleichung 4). Die Reduktion getesteter Eigenwerte impliziert eine verringerte Signalstärke der Eigenwerte, die mit Faktoren korrespondieren. Dadurch kann eine erhöhte Wahrscheinlichkeit von Unterschätzungen der optimalen Faktorenzahl durch NEST in Anwesenheit substantieller Kreuzladungen vorhergesagt werden. Diese Vorhersage bedarf jedoch simulationsbasierter Untersuchungen: Alleine aus der Feststellung, dass getestete Eigenwerte durch Kreuzladungen reduziert werden, ergibt sich nicht, ob und wie häufig es dadurch zu Unterschätzungen durch NEST kommt und wie sich NEST in Anwesenheit substantieller Kreuzladungen mit anderen Methoden vergleicht.

Exploratory Graph Analysis

Golino und Epskamp (2017) veröffentlichten mit EGA eine Methode zur Bestimmung der Faktorenzahl, die sich nicht in eine etablierte Klasse von Methoden einordnen lässt, sondern einen gänzlich neuen Ansatz zur Bestimmung der

Faktorenzahl darstellt. EGA basiert auf *psychometrischen Netzwerkmodellen*, die eine Alternative zu Faktorenmodellen bei der theoretischen Erklärung beobachteter Korrelationen sind. Netzwerkmodelle unterscheiden sich darin von Faktorenmodellen, dass Korrelationen gemessener Variablen nicht durch latente Faktoren erklärt werden, sondern durch unmittelbare paarweise Abhängigkeiten der gemessenen Variablen (Cramer et al., 2010, 2012; Epskamp et al., 2017, 2018; Marsman et al., 2018). Die Motivation, unmittelbare Abhängigkeiten gegenüber gemeinsamen Faktoren als kausale Erklärung für Korrelationen zu bevorzugen, wurde für Anwendungen im Bereich der Psychopathologie vielfach theoretisch begründet (Blanken et al., 2018; Borsboom, 2017; Borsboom & Cramer, 2013; Borsboom et al., 2011; Fried & Cramer, 2017; Isvoranu et al., 2021). Die inhaltliche Kernaussage psychometrischer Netzwerkmodelle komorbider Symptome ist, dass sich Symptome gegenseitig bedingen und nicht durch einen hypothetischen latenten Faktor bedingt seien[6].

Trotz der inhaltlichen Unterschiede zwischen psychometrischen Netzwerkmodellen und Faktorenmodellen bestimmt EGA die Faktorenzahl für einen Datensatz mithilfe eines Netzwerkmodels derselben Daten. Die angenommene Korrespondenz zwischen Netzwerk- und Faktorenmodellen offenbart sich in einer graphischen Repräsentation der Netzwerkmodelle (siehe Golino & Epskamp, 2017). Ein Netzwerkmodell lässt sich als Graph veranschaulichen, dessen Knotenmenge die Menge gemessener Variablen repräsentiert und dessen Kantenmenge die paarweisen Abhängigkeiten zwischen Variablen repräsentiert. Golino und Epskamp (2017) zufolge kann die Beobachtung dicht vernetzter Knoten in Netzwerkmodellgraphen mit dem Einfluss latenter Faktoren gleichgesetzt werden. EGA bestimmt daher die Faktorenzahl für einen Datensatz, indem ein Netzwerkmodellgraph für den Datensatz zunächst berechnet und anschließend in Teilgraphen partitioniert wird. Zur Bestimmung der Faktorenzahl wird in EGA eine Partitionierung gesucht, in der Knoten innerhalb von Teilgraphen dicht vernetzt und Knoten zwischen Teilgraphen möglichst unverbunden sind. Die Betrachtung, dass dichte Teilgraphen in Netzwerkmodellgraphen mögliche Faktoren anzeigen, ist in der Literatur zu psychometrischen Netzwerkmodellen verbreitet (Cramer et al., 2012; Epskamp et al., 2018; van der Maas et al., 2006). Konkret bestimmt EGA die Faktorenzahl als

[6] Unmittelbare paarweise Abhängigkeiten werden in Netzwerkmodellen berechnet anhand der invertierten Stichproben-Varianz-Kovarianzmatrix, deren Einträge nach Standardisierung die Partialkorrelation zweier gemessener Variablen konditional zu allen weiteren gemessenen Variablen angeben (für eine technische Einführung psychometrischer Netzwerkmodelle, siehe Williams & Rast, 2020).

äquivalent zur Anzahl der Teilgraphen in einer algorithmisch ermittelten Partitionierung des Netzwerkmodellgraphen[7].

Golino et al. (2020) untersuchten EGA unter Verwendung des *Walktrap*-Algorithmus (Pons & Latapy, 2006) zur Partitionierung von Netzwerkmodellgraphen. Als Alternative zu Walktrap wurde in einer weiteren Arbeit (Christensen et al., 2020) der *Louvain*-Algorithmus (Blondel et al., 2008) zur Partitionierung diskutiert. Der Anlass, EGA für die vorliegende Dissertation insbesondere im Hinblick auf Kreuzladungen erneut zu untersuchen, ergibt sich aus einer kritischen Restriktion der verwendeten Algorithmen zur Partitionierung: In Partitionierungen durch Walktrap und Louvain sind die Knotenmengen der ermittelten Teilgraphen disjunkt, sodass die Teilgraphen als nicht-überlappend beschrieben werden können. Nicht-überlappende Teilgraphen limitieren jedoch die Korrespondenz zwischen Netzwerkmodellen und Faktorenmodellen. In EGA-Lösungen sind Variablen notwendigerweise jeweils nur mit einem Faktor assoziiert, da Knoten nur Element der Knotenmenge eines Teilgraphen sein können. Diese Restriktion in EGA-Lösungen ist nicht mit Faktorenmodellen vereinbar, in denen Variablen durch substantielle Kreuzladungen mit mehreren Faktoren assoziiert sind.

Golino et al. (2020) wiesen bereits darauf hin, dass die Validierung von EGA weitere Simulationsstudien erfordert, um die Anwendbarkeit von EGA in Anwesenheit substantieller Kreuzladungen zu untersuchen. Für die vorliegende Dissertation wurde untersucht, ob und wie häufig EGA-Lösungen von der optimalen Faktorenzahl bei substantiellen Kreuzladungen abweichen. In allen durchgeführten Simulationen wurden dabei zwei EGA-Implementierungen für jeden simulierten Datensatz angewendet: eine Implementierung mit Walktrap ($EGA_{Walktrap}$) und eine Implementierung mit Louvain ($EGA_{Louvain}$) zur Partitionierung der Netzwerkmodellgraphen.

Simulationen 1a und 1b

Die Simulationen 1a und 1b stellen eine experimentelle Untersuchung der Robustheit von NEST und EGA gegen Kreuzladungen dar. In Simulation 1a wurden Datensätze ohne Kreuzladungen und Simulation 1b Datensätze mit Kreuzladungen simuliert, während alle weiteren Datenmerkmale zwischen den Simulationen konstant blieben.

[7] Die Literatur zu psychometrischen Netzwerkmodellen und EGA enthält keine konsistente Theorie, in der Kriterien für die Partitionierung eines Netzwerkmodellgraphen in Übereinstimmung mit einem äquivalenten Faktorenmodell definiert werden.

Die Lösungen von NEST und EGA wurden außerdem mit Lösungen von PA verglichen, um vorherige Vergleiche von NEST bzw. EGA mit PA um Bedingungen mit substantiellen Kreuzladungen zu erweitern[8].

Die Simulationen 1a und 1b basierten auf einem gemeinsamen vollständig-gekreuzten Design. Nach dem gemeinsamen Design wurden Datenmerkmale manipuliert – unter anderem die Kommunalität gemessener Variablen – zur Simulation heterogener Anwendungsfälle (für eine vollständige Auflistung der manipulierten Datenmerkmale und deren Faktorstufen, siehe Brandenburg & Papenberg, 2022, S. 5). Die Merkmalskombinationen implizierten Faktorenmodelle, die wiederum die Populationskorrelationsmatrizen multivariater Normalverteilungen implizierten (siehe Gleichung 4), aus denen die simulierten Datensätze gezogen wurden. Jeder Datensatz wurde durch alle untersuchten Methoden ausgewertet und alle Methodenlösungen wurden gespeichert.

In Simulation 1a wurde die spezifizierte Kommunalität aller simulierten Variablen jeweils durch eine Faktorladung der Variablen auf einen zugewiesenen Faktor determiniert. Der alleinige Unterschied in Simulation 1b war, dass die Kommunalität durch Kreuzladungen auf den gleichen zugewiesenen Faktor wie in Simulation 1a und einen weiteren, zufällig ausgewählten Faktor determiniert wurde[9]. Die jeweiligen Anteile der beiden Faktoren an der Kommunalität wurden in Simulation 1b unabhängig für jede Variable zufällig zwischen 0 und 100 % bestimmt (für eine technische Beschreibung der Simulationen 1a und 1b, siehe Brandenburg & Papenberg, 2022, S. 6).

Die individuelle Lösung jeder Methode bei der Bestimmung der Faktorenzahl wurde für alle simulierten Datensätze in eine von vier Kategorien eingeordnet. Lösungen wurden als *akkurat* bewertet, wenn die Methode die Faktorenzahl des Faktorenmodells reproduzierte, unter dem der jeweilige Datensatz simuliert worden war. Andernfalls wurden Lösungen als *unterschätzt* bzw. *überschätzt* bewertet, wenn

[8] Der vorliegende Ergebnisbericht beschränkt sich auf eine PA-Implementierung. Bei der Durchführung der Simulationen 1a bis 2f wurden insgesamt sechs PA-Implementierungen parallel zu NEST und EGA verwendet. Detailliertere Ausführungen zu allen PA-Implementierungen finden sich in der ersten Einzelarbeit (Brandenburg & Papenberg, 2022, S. 7) Die PA-Implementierung im vorliegenden Bericht entspricht der Implementierung $PA_{FA\text{-}SMC\text{-}50}$ der ersten Einzelarbeit. Die PA-Implementierung im vorliegenden Bericht ist repräsentativ für die Schlussfolgerungen, die bezüglich PA im Vergleich zu NEST und EGA gemacht werden können.

[9] Die Simulationen 1a und 1b hätten äquivalent als eine einzelne Simulation beschrieben werden können, in der die Zahl der Faktorladungen pro Variable als unabhängige Variable manipuliert worden wäre. In der vorliegenden Dissertation wird stattdessen zwischen den Simulationen 1a und 1b unterschieden, um Konsistenz mit der ersten Einzelarbeit (Brandenburg & Papenberg, 2022) zu wahren.

sie die spezifizierte Faktorenzahl unter- bzw. überschritten. Außerdem wurden Lösungen als *undefiniert* bewertet, wenn die jeweilige Methode für den Datensatz keine Faktorenzahl bestimmen konnte. Im Nachfolgenden wird die *Genauigkeit* einer Methode als Anteil der akkuraten Lösungen an allen vier Kategorien definiert.

Ergebnisse der Simulationen 1a und 1b

In Tabelle 1 sind die Ergebnisse der Simulationen 1a und 1b zusammengefasst. Die Genauigkeit der Methoden für Datensätze ohne Kreuzladungen in Simulation 1a unterschied sich nur um wenige Prozentpunkte. Ohne Kreuzladungen war die Genauigkeit von NEST und der EGA-Implementierungen ($EGA_{Walktrap}$ und $EGA_{Louvain}$) vergleichbar mit PA. Kreuzladungen reduzierten jedoch die Genauigkeit aller Methoden in Simulation 1b. Für NEST erhöhten Kreuzladungen die Häufigkeit von Unterschätzungen konsistent mit der Überlegung bezüglich verringerter Eigenwerte (siehe oben)[10]. Für $EGA_{Walktrap}$ und $EGA_{Louvain}$ reduzierten Kreuzladungen dagegen die Häufigkeit von sowohl Unter- als auch Überschätzungen erhöhten.

Tabelle 1

Lösungen in den Simulationen 1a und 1b

Simulation	Lösung	Methode			
		NEST	$EGA_{Walktrap}$	$EGA_{Louvain}$	PA
Simulation 1a (153,600 Datensätze)	überschätzt	1.5	1.2	1.4	8.7
	akkurat	84.2	79.2	79.8	83.3
	unterschätzt	14.0	8.0	9.3	8.0
	undefiniert	0.2	11.6	9.5	0
Simulation 1b (153,600 Datensätze)	überschätzt	0.8	7.9	10.5	4.4
	akkurat	49.5	21.1	24.4	53.7
	unterschätzt	41.2	53.6	50.0	41.7
	undefiniert	8.5	17.4	15.1	0.2

Bemerkung: Die Werte geben den prozentualen Anteil der jeweiligen Lösungskategorie an allen Lösungen der entsprechenden Methode in der entsprechenden Simulation an.

Kreuzladungen reduzierten die Genauigkeit der EGA-Implementierungen stärker als die Genauigkeit der anderen Methoden. NEST war in Simulation 1b weiterhin ähnlich genau wie PA und daher ähnlich robust gegen Kreuzladungen. Das auffällige Genauigkeitsdefizit beider EGA-Implementierungen durch Kreuzladungen deutet darauf hin, dass die Bestimmung der Faktorenzahl durch nicht-überlappende Teilgraphen nicht mit den spezifizierten Kreuzladungen in

[10] Die Zunahme undefinierter NEST-Lösungen ist durch technische Aspekte der simulierten Datensätze erklärbar und wird in der ersten Einzelarbeit (Brandenburg & Papenberg, 2022, S. 8) diskutiert.

Simulation 1b kompatibel war. Die Inkompatibilität nicht-überlappender Teilgraphen mit Kreuzladungen wird dadurch untermauert, dass der Genauigkeitsverlust von EGA nicht nur durch undefinierte Lösungen entstand, sondern größtenteils durch Lösungen, in denen die Anzahl der Teilgraphen nicht der optimalen Faktorenzahl entsprach.

Simulationen 2a bis 2f

In den Simulationen 1a und 1b wurden Effekte von Kreuzladungen auf die Genauigkeit der untersuchten Methoden bei isolierter Manipulation nachgewiesen. Darauf aufbauend folgten die Simulationen 2a bis 2f der Zielsetzung, den Vergleich der Methoden um weitere Simulationsbedingungen mit Kreuzladungen unter Berücksichtigung ökologischer Validität zu erweitern. Dazu wurden in den Simulationen 2a bis 2f Datensätze unter Faktorenmodellen mit Kreuzladungen simuliert, die – im Gegensatz zu den Faktorenmodellen in Simulation 1b – in empirischen Anwendungen diskutiert werden.

Bei den spezifizierten Faktorenmodellen der Simulationen 2a bis 2f handelte es sich um sogenannte Circumplexmodelle (Guttman, 1954b). Grundsätzlich sind Circumplexmodelle ebenfalls Modelle für Korrelationen, die zwar nicht durch Faktorenanalysen geschätzt werden müssen (siehe Browne, 1992), aber als Faktorenmodelle betrachtet werden können (Acton & Revelle, 2002, 2004; Fabrigar et al., 1997; Remmington et al., 2000). Circumplexmodelle enthalten in der Betrachtung als Faktorenmodelle zwei unabhängige Faktoren. Das charakteristische Merkmal von Circumplexmodellen ist dabei, dass die Zeilen der Faktorladungsmatrix Λ als kartesische Koordinaten gelesen werden können, die sich kreisförmig um den Ursprung anordnen (siehe Brandenburg & Papenberg, 2022, S. 10).

Die inhaltliche Kernaussage in Circumplexmodellen ist, dass die untersuchte Domäne zwei wesentliche Faktoren enthält und dass messbare Variablen dieser Domäne in allen Orientierungen relativ zu diesen Faktoren existieren (Acton & Revelle, 2004; Gurtman & Pincus, 2003). Anwendungen von Circumplexmodellen finden sich in der Literatur beispielsweise für Variablen mit Bezug zu Affekt (Fabrigar et al., 1997; Gurtman & Pincus, 2000, 2003; Remmington et al., 2000; Tracey, 2000) und Interpersönlichkeitspsychologie (Alden et al., 1990; Bordeaux et al., 2018; Di Blas et al., 2012; Hopwood et al., 2011; Hopwood & Good, 2019; Locke, 2000; Zimmermann & Wright, 2017).

Das Faktorenzahlproblem wird im Kontext von Circumplexmodellen nicht prominent diskutiert. Dennoch stellen Circumplexmodelle eine nützliche Instanz des Faktorenzahlproblems dar, die zur Validierung von Methoden zur Bestimmung der Faktorenzahl in Anwesenheit substantieller Kreuzladungen beitragen kann. Zum einen implizieren Circumplexmodelle eine eindeutige Faktorenzahl. Es ist bisher nicht untersucht worden, ob NEST bzw. EGA die optimale Faktorenzahl auch dann zuverlässig anzeigen kann, wenn die Faktoren in Form des charakteristischen kreisförmigen Musters der Faktorladungen angezeigt werden. Zum anderen impliziert das kreisförmige Muster der Faktorladungen, dass die gemessenen Variablen Kreuzladungen zeigen und mit mehr als einem Faktor assoziiert sind.

In den Simulationen 2a bis 2f wurden die Lösungen von NEST, EGA und PA unter Berücksichtigung verschiedener Definitionen und Anwendungsformen von Circumplexmodellen untersucht. Neben den zwei Faktoren, deren Faktorladungen das charakteristische kreisförmige Muster kennzeichnen, enthalten Circumplexmodelle teilweise einen zusätzlichen Generalfaktor, auf den alle gemessenen Variablen laden und der eine generelle Tendenz zu hohen bzw. niedrigen Variablenwerten beschreibt (Alden et al., 1990; Hopwood et al., 2011; Rogoza et al., 2021; Tracey, 2000; Zimmermann & Wright, 2017). Außerdem finden sich in der Literatur Anwendungen von Circumplexmodellen für gemessene Variablen (Alden et al., 1990; Bordeaux et al., 2018; Di Blas et al., 2012; Hopwood et al., 2011) und für aggregierte Werte mehrerer gemessener Variablen (Gurtman & Pincus, 2000; Hopwood & Good, 2019; Rogoza et al., 2021; Zimmermann & Wright, 2017). Im Nachfolgenden bezeichnen *Circumplexmodelle auf Variablenebene* den Fall, dass einzelne gemessene Variablen das kreisförmige Ladungsmuster zeigten. *Circumplexmodelle auf Subskalenebene* bezeichnen hingegen den Fall, dass die Variablen auf Subskalen aufgeteilt wurden und die Summenwerte aller Variablen in den jeweiligen Subskalen das kreisförmige Ladungsmuster zeigten.

Die nachfolgenden Abschnitte fassen zusammen, wie sich die spezifizierten Circumplexmodelle zwischen den Simulationen 2a bis 2f unterschieden[11]. In allen Simulationen implizierten die Circumplexmodelle Populationskorrelationsmatrizen multivariater Normalverteilungen, aus denen die simulierten Datensätze gezogen

[11] Die erste Einzelarbeit (Brandenburg & Papenberg, 2022) enthält Illustrationen der spezifizierten Faktorladungen, der modellimplizierten Korrelationen und der daraus resultierenden Eigenwerte für die Circumplexmodelle der Simulationen 2a bis 2f. Ein Unterschied der Kreuzladungen in Circumplexmodellen zu den Kreuzladungen in Simulation 1b war, dass in den Circumplexmodellen explizit negative Faktorladungen vorgesehen waren (siehe Remmington et al., 2000; Rogoza et al., 2021; Zimmermann & Wright, 2017), die keine Verringerung von Eigenwerten implizierten.

wurden. Dabei wurden in den Simulationen 2a bis 2f erneut verschiedene Datenmerkmale zur Simulation heterogener Anwendungsfälle manipuliert (für die vollständigen Designs der Simulationen 2a bis 2f sowie eine technische Beschreibung der Spezifikation der Modellparameter, siehe Brandenburg & Papenberg, 2022, S. 10-14). In den Simulationen 2a bis 2f wurden dieselben Implementierungen von NEST, EGA und PA wie in den Simulationen 1a und 1b verwendet und die Methodenlösungen wurden weiterhin nach den vier Kategorien (akkurat, unterschätzt, überschätzt, undefiniert) bewertet.

In Simulation 2a wurden Circumplexmodelle auf Variablenebene spezifiziert, in denen gemessene Variablen lediglich auf die zwei für das kreisförmige Ladungsmuster notwendigen Faktoren luden. Die optimale Faktorenzahl wurde in Simulation 2a dementsprechend auf zwei festgelegt.

In Simulation 2b wurden Circumplexmodelle auf Variablenebene spezifiziert, in denen gemessene Variablen ebenfalls kreisförmig auf zwei Faktoren luden und zusätzlich auf einen Generalfaktor. Aufgrund des Generalfaktors wurde die optimale Faktorenzahl in Simulation 2b auf drei festgelegt.

In Simulation 2c wurden Circumplexmodelle auf Subskalenebene spezifiziert. Der Modellterm $\Lambda\Psi\Lambda^T$ (siehe Gleichung 4) setzte sich in in Simulation 2c aus zwei Aspekten zusammen: In der Faktorladungsmatrix Λ der gemessenen Variablen wurde für jede Variable eine Faktorladung auf einen von insgesamt acht Faktoren spezifiziert, sodass die gemessenen Variablen acht distinkte eindimensionale Subskalen bildeten (siehe Brandenburg & Papenberg, 2022, S. 10-11, für eine Erläuterung der Festlegung auf acht Subskalen). Die Interfaktorenkorrelationsmatrix Ψ wurde durch ein höhergeordnetes Circumplexmodell auf Subskalenebene determiniert, in dem die acht Faktoren der Variablenebene ein kreisförmiges Ladungsmuster auf zwei Faktoren höherer Ordnung zeigten. Die Modelle in Simulation 2c wurden demnach so spezifiziert, dass die einzelnen simulierten Variablen Subskalen gemäß Λ bildeten, die wiederum gemäß Ψ korrelierten. Aggregationen von Subskalen wurden in Simulation 2c nicht berechnet; die untersuchten Methoden bestimmten die Faktorenzahl für die einzelnen gemessenen Variablen. Da die gemessenen Variablen in Simulation 2c jeweils einen von acht Faktoren auf Variablenebene anzeigten und nur die Interfaktorenkorrelationen durch Circumplexmodelle determiniert wurden, wurde die optimale Faktorenzahl bei Auswertung der gemessenen Variablen in Simulation 2c auf acht festgelegt.

In Simulation 2d wurden ebenfalls Circumplexmodelle auf Subskalenebene spezifiziert. Dazu wurde das Vorgehen aus Simulation 2c übernommen, jedoch mit dem Unterschied, dass die Circumplexmodelle auf Subskalenebene für Ψ in Simulation 2d einen zusätzlichen Generalfaktor enthielten und nach der Methode aus Simulation 2b spezifiziert wurden. Die simulierten Datensätze zur Untersuchung der Methoden enthielten in Simulation 2d erneut die einzelnen gemessenen Variablen und keine Aggregationen. Daher wurde die optimale Faktorenzahl in Simulation 2d ebenfalls auf acht festgelegt.

In Simulation 2e wurden die simulierten Datensätze aus Simulationen 2c wiederverwendet und in Simulation 2f wurden die simulierten Datensätze aus Simulation 2d wiederverwendet. Im Gegensatz zu den vorherigen Simulationen wurden in den Simulationen 2e und 2f die Summenwerte der acht Subskalen für jeden Datensatz berechnet. Die untersuchten Methoden sollten die optimale Faktorenzahl für die acht Summenwerte jedes Datensatzes bestimmen. Da die Korrelationen der Summenwerte durch die Circumplexmodelle auf Subskalenebene der Simulationen 2c und 2d determiniert wurden, wurde die optimale Faktorenzahl in Simulation 2e auf zwei und in Simulation 2f auf drei festgelegt.

Ergebnisse der Simulationen 2a bis 2f

In Tabelle 2 sind die Ergebnisse für die untersuchten Methoden in den Simulationen 2a bis 2f zusammengefasst. In den Simulationen 2a und 2b reproduzierten NEST und PA die optimale Faktorenzahl der Circumplexmodelle auf Variablenebene für nahezu alle Datensätze. Im Kontrast dazu reproduzierten beide EGA-Implementierungen ($EGA_{Walktrap}$ und $EGA_{Louvain}$) die optimale Faktorenzahl deutlich seltener und neigten zu Überschätzungen. Das Genauigkeitsdefizit von EGA legt nahe, dass speziell EGA ungeeignet war, die optimale Faktorenzahl für Datensätze zu bestimmen, in denen die Faktoren in Form des charakteristischen kreisförmigen Ladungsmusters angezeigt wurden.

Tabelle 2

Lösungen in den Simulationen 2a bis 2f

Simulation	Lösung	Methode			
		NEST	$EGA_{Walktrap}$	$EGA_{Louvain}$	PA
Simulation 2a (9,600 Datensätze)	überschätzt	2.0	73.6	77.7	3.2
	akkurat	98.0	6.9	2.9	96.8
	unterschätzt	0	0.1	2.6	0
	undefiniert	0	19.3	16.9	0
Simulation 2b (28,800 Datensätze)	überschätzt	2.8	62.7	65.8	6.1
	akkurat	96.5	16.9	14.0	93.8
	unterschätzt	0.7	1.3	3.8	0.1
	undefiniert	0	19.0	16.4	0
Simulation 2c (76,800 Datensätze)	überschätzt	6.2	1.1	0.5	12.1
	akkurat	70.9	66.2	65.0	67.3
	unterschätzt	22.7	18.2	22.0	20.7
	undefiniert	0.3	14.5	12.5	0
Simulation 2d (76,800 Datensätze)	überschätzt	6.9	1.1	0.5	13.0
	akkurat	71.6	67.3	66.7	68.6
	unterschätzt	21.2	18.3	21.6	18.3
	undefiniert	0.3	13.2	11.3	0
Simulation 2e (76,800 Datensätze)	überschätzt	0.7	10.7	15.9	12.2
	akkurat	89.7	67.3	62.2	85.8
	unterschätzt	9.2	0.9	5.9	2.0
	undefiniert	0.4	21.1	15.9	0
Simulation 2f (76,800 Datensätze)	überschätzt	0.1	3.0	5.9	11.0
	akkurat	72.1	42.9	51.9	82.5
	unterschätzt	21.9	28.3	22.9	6.5
	undefiniert	5.8	25.9	19.2	0

Bemerkung: Die Werte geben den prozentualen Anteil der jeweiligen Lösungskategorie an allen Lösungen der entsprechenden Methode in der entsprechenden Simulation an.

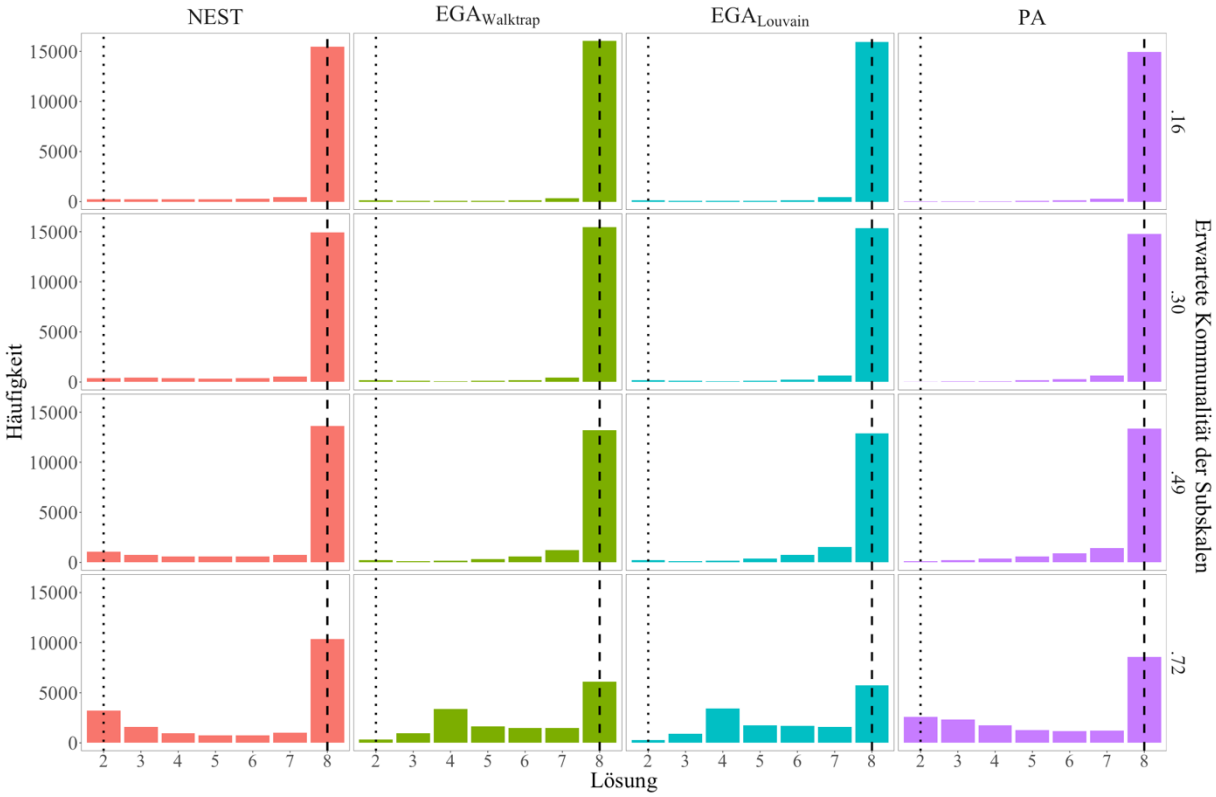
In den Simulationen 2c und 2d unterschied sich die Genauigkeit aller Methoden um wenige Prozentpunkte. Der geringe Unterschied zwischen NEST und EGA für Circumplexmodelle auf Subskalenebene bei eindimensionalen Subskalen ist konsistent mit den Ergebnissen der Simulation 1a; in beiden Fällen wurden die beobachteten Korrelationen primär ohne Kreuzladungen determiniert. Eine detailliertere Betrachtung der Lösungen als Funktion manipulierter Kommunalität der Subskalen in den Circumplexmodellen[12] auf Subskalenebene zeigt, wie sich Lösungen mit zunehmendem Einfluss der Circumplexmodelle veränderten. Je höher die Kommunalität der Subskalen war, desto größer war der Einfluss der

[12] Die Kommunalität der Subskalen in Circumplexmodellen auf Subskalenebene in den Simulationen 2c und 2d wurden gemeinsam mit anderen Datenmerkmalen in einem vollständig-gekreuzten Simulationsdesign manipuliert. Für jede Subskala eines Datensatzes wurde die Kommunalität im Circumplexmodell unabhängig und identisch verteilt aus vordefinierten Gleichverteilungen (parallel zu Golino et al., 2020) gezogen: $U(.40^2 - 0.05, .40^2 + 0.05)$, $U(.55^2 - 0.05, .55^2 + 0.05)$, $U(.70^2 - 0.05, .70^2 + 0.05)$, $U(.85^2 - 0.05, .85^2 + 0.05)$.

Circumplexmodelle auf die beobachteten Korrelationen. Abbildung 1 zeigt, dass NEST und PA in Simulation 2c die optimale Faktorenzahl auf Variablenebene (acht Faktoren) mit zunehmender Kommunalität der Subskalen häufiger unterschätzten und dabei zunehmend zur Faktorenzahl der Circumplexmodelle (zwei Faktoren) tendierten. Die EGA-Implementierungen unterschätzten die Faktorenzahl ebenfalls häufiger mit zunehmender Kommunalität der Subskalen, tendierten dabei aber zu einer höheren Faktorenzahl als die der Circumplexmodelle. Die Divergenz der EGA-Lösungen von der Faktorenzahl der Circumplexmodelle durch EGA in Simulation 2c stimmt mit den Überschätzungen der optimalen Faktorenzahl durch EGA in den Simulationen 2a und 2b überein. In Simulation 2d zeigte sich das gleiche Muster mit zunehmender Kommunalität der Subskalen wie in Simulation 2c: NEST und PA neigten bei maximaler Kommunalität der Subskalen zur Faktorenzahl der Circumplexmodelle auf Subskalenebene; EGA neigte stattdessen zu einer höheren Faktorenzahl (siehe Brandenburg & Papenberg, 2022, S. 19, für eine Illustration dieses Ergebnismusters in Simulation 2d analog zu Abbildung 1).

Abbildung 1

Lösungen in Simulation 2c als Funktion der erwarteten Kommunalität der Subskalen



Bemerkung: Die Säulen zeigen die absoluten Häufigkeiten ausgewählter Lösungen (zwei bis acht Faktoren) der jeweiligen Methode in Simulation 2c getrennt nach der manipulierten Kommunalität der Subskalen in den spezifizierten Circumplexmodellen auf Subskalenebene. Von oben nach unten repräsentieren die Zeilen aufsteigenden Einfluss der Circumplexmodelle auf die simulierten Datensätze. Die gestrichelte Linie zeigt die optimale Faktorenzahl in Simulation 2c (acht Faktoren) an. Die gepunktete Linie zeigt die Faktorenzahl der Circumplexmodelle auf Subskalenebene (zwei Faktoren) an.

Die Ergebnisse der Simulationen 2e und 2f betreffen die Lösungen der untersuchten Methoden, wenn die Faktorenzahl für die Summenwerte der Subskalen der simulierten Datensätze aus den Simulationen 2c und 2d bestimmt werden sollte. NEST und PA waren in Simulation 2e bei der Auswertung von Summenwerten genauer als in Simulation 2c, da sie die die optimale Faktorenzahl seltener unterschätzten. In Simulation 2f unterschätzte NEST die optimale Faktorenzahl häufiger als PA und war insgesamt ungenauer als PA. Für EGA stimmen die Ergebnisse der Simulationen 2e und 2f mit den restlichen Simulationen darin überein, dass beide EGA-Implementierungen ungenauer waren als NEST und PA für Datensätze, in denen Circumplexmodelle die beobachteten Korrelationen determinierten.

Diskussion der Simulationen 1a bis 2f

Die Simulationen 1a bis 2f erweitern die bisherige Studienlage zur Validierung von NEST (Achim, 2017) und EGA (Golino et al., 2020) um einen direkten Vergleich der beiden Methoden und um einen Fokus auf substantielle Kreuzladungen.

In den Simulationen 1a und 1b wurde die Anzahl von Faktorladungen pro Variable isoliert manipuliert, um die Auswirkungen von Kreuzladungen auf die Lösungen von NEST und EGA zu untersuchen. Die Ergebnisse der Simulation 1a bestätigen die vorherigen Studien (Achim, 2017; Golino et al., 2020) insofern, als NEST und EGA ohne Kreuzladungen zumindest vergleichbar genau wie PA waren. Die Ergebnisse der Simulation 1b zeigen darüber hinaus, dass die Manipulation von Kreuzladungen – in der Form von Simulation 1b – die Genauigkeit von NEST und PA in ähnlichem Ausmaß reduzierte. Die vorliegenden Ergebnisse erweitern daher die Validierung von NEST im Vergleich zu PA um Bedingungen mit substantiellen Kreuzladungen.

Außerdem zeigen die Ergebnisse der Simulation 1b, dass Kreuzladungen die Genauigkeit von EGA stärker reduzierten als die Genauigkeit von NEST. Die stärkere Robustheit von NEST und PA gegen Kreuzladungen im Vergleich zu EGA zeigt einen zuvor unentdeckten Nachteil für EGA. Das Genauigkeitsdefizit von EGA im Vergleich zu anderen Methoden weist darauf hin, dass die bisherige Validierung von EGA (Golino et al., 2020) nicht auf Simulationsbedingungen mit substantiellen Kreuzladungen zutrifft.

In den Simulationen 2a bis 2f wurden Circumplexmodelle als Instanz des Faktorenzahlproblems mit substantiellen Kreuzladungen spezifiziert. Die Ergebnisse zu NEST stimmen mit den Simulationen 1a und 1b darin überein, dass NEST die optimale Faktorenzahl erneut größtenteils vergleichbar genau wie PA reproduzierte. Die nahezu perfekte Genauigkeit von NEST in den Simulationen 2a und 2b erweitert die Validierung von NEST im Kontext substantieller Kreuzladungen über Simulation 1b hinaus, da die Circumplexmodelle zwar Kreuzladungen implizierten, aber die Genauigkeit von NEST nicht reduzierten.

Zu EGA wurde in den Simulationen 2a bis 2f nachgewiesen, dass das Genauigkeitsdefizit von EGA bei substantiellen Kreuzladungen auch bei Circumplexmodellen auftrat. Das Genauigkeitsdefizit von EGA war insbesondere in den Simulationen 2a und 2b größer als in Simulation 1b. Die Ergebnisse der Simulationen 2a bis 2f weisen gemeinsam mit Simulation 1b auf eine Limitation von

EGA hin: Die Verwendung von EGA zur Bestimmung der Faktorenzahl anstelle von NEST oder PA setzt die Annahme voraus, dass keine substantiellen Kreuzladungen vorliegen. Das Ausmaß, bis zu welchem Grad EGA robust gegen Kreuzladungen sein könnte, lässt sich jedoch anhand der vorliegenden Ergebnisse nicht präzisieren, da die Anzahl der kreuzladenden Variablen nicht in Abstufungen manipuliert wurde (Beispiel: Li et al., 2020).

Wie eingangs formuliert ergibt sich aus theoretischen Überlegungen, dass nicht-überlappende Teilgraphen eines Netzwerkmodellgraphen in EGA nicht mit Faktoren kompatibel sind, wenn Faktoren substantielle Kreuzladungen in den gemessenen Variablen aufweisen. Diese Inkompatibilität zwischen nicht-überlappenden Teilgraphen und Faktoren wird durch die niedrige Genauigkeit von EGA in den durchgeführten Simulationen anhand von Daten bestätigt. Die Inkompatibilität wird zusätzlich dadurch untermauert, dass sie mit zwei verschiedenen Algorithmen zur Partitionierung von Netzwerkmodellgraphen (Walktrap und Louvain) beobachtet wurde.

Der Nachweis kreuzladungsbedingter Inkompatibilität nicht-überlappender Teilgraphen und Faktoren limitiert die Position, dass Teilgraphen mit dicht-verbundenen Knoten den möglichen Einfluss von Faktoren anzeigen würden (Cramer et al., 2012; Epskamp et al., 2018; Golino & Epskamp, 2017; Golino et al., 2020; van der Maas et al., 2006): Bei substantiellen Kreuzladungen und den hier untersuchten Algorithmen zur Detektion von Teilgraphen scheint diese Korrespondenz nicht zu gelten. Inwiefern die Inkompatibilität nicht-überlappender Teilgraphen und Faktoren auf andere Algorithmen zur Partitionierung von Netzwerkmodellgraphen zutrifft, bedarf jedoch weiterer Forschung – insbesondere zu Algorithmen, die überlappende Teilgraphen zulassen (Beispiel: Blanken et al., 2018).

Die vorliegenden Ergebnisse zu EGA betreffen keine Anwendungen psychometrischer Netzwerkmodelle in Kontexten, in denen paarweise Abhängigkeiten der Variablen als Ursache für Korrelationen gesichert feststehen. Es wird jedoch aus methodischer Perspektive kritisch diskutiert, unter welchen Bedingungen der theoretisch motivierte Verzicht auf Faktoren zur Erklärung von Korrelationen (Borsboom et al., 2011; Epskamp et al., 2018; Isvoranu et al., 2021) gerechtfertigt ist. Netzwerkmodelle reflektieren paarweise Abhängigkeiten nur unter Kontrolle gemessener Variablen und nicht unter Kontrolle latenter Variablen (Hallquist et al., 2021; Neal & Neal, 2021). Hallquist et al. (2021) diskutieren daher

den Umgang mit Faktoren, die per Definition latent und damit in Netzwerkmodellen unkontrolliert sind. Die Interpretation von Netzwerkmodellen erfordere demnach die Annahme, dass Faktoren die beobachteten Korrelationen nicht erklären können, oder eine Diskussion, an welchen Stellen in Netzwerkmodellen Faktoren als konfundierte Erklärung für Korrelationen in Frage kommen. Obwohl sich die Entwicklung von EGA (siehe Golino & Epskamp, 2017) nicht auf Hallquist et al. beziehen konnte, legt die bisher angenommene Korrespondenz zwischen Teilgraphen und Faktoren nahe, dass EGA zur Erkennung von Faktoren in Netzwerkmodellen beitragen könne. Die vorliegenden Ergebnisse untermauern jedoch, dass EGA nur unter der – für Faktorenmodelle restriktiven – Zusatzannahme Faktoren zuverlässig anzeigen kann, dass keine substantiellen Kreuzladungen vorliegen. Es bedarf daher weiterer Forschung zu Methoden, die zur Diskussion hypothetischer Faktoren in Netzwerkmodellen beitragen, da EGA als naheliegender Ansatz fundamentale Probleme aufweist.

Zusammengefasst unterstützt der Methodenvergleich in der vorliegenden Dissertation eine Präferenz von NEST[13] gegenüber EGA zur Bestimmung der Faktorenzahl. Die Ergebnisse der Simulationen 1a bis 2f bestätigen die theoretische Überlegung, dass EGA nicht mit Kreuzladungen kompatibel ist. Parallel dazu belegen die Ergebnisse der Simulationen 1a bis 2f, dass NEST im Vergleich zu EGA bei substantiellen Kreuzladungen weniger restriktiv und genauer bei der Bestimmung der optimalen Faktorenzahl ist. Im weiteren Verlauf der vorliegenden Dissertation werden die Validierung von NEST und der Vergleich zwischen NEST und PA mit zusätzlichen Problemstellungen fortgesetzt. Basierend auf den Ergebnissen der Simulationen 1a bis 2f wurde EGA in weiteren Simulationen nicht berücksichtigt.

Kategoriale Variablen

Die vorangehenden Simulationsergebnisse zeigen, dass NEST anhand von Stichprobenkorrelationsmatrizen die optimale Faktorenzahl im Vergleich zu weiteren Methoden mit hoher Genauigkeit reproduzieren konnte. Die zentrale Fragestellung im weiteren Verlauf der vorliegenden Dissertation betrifft die

[13] Als simulationsbasierter Vergleich beruhen die hier diskutierten Argumente pro NEST primär auf der Genauigkeit von NEST im Vergleich zu EGA bzw. PA. Zusätzliche theoretische Vorteile von NEST gegenüber PA (siehe Achim, 2017) werden in der ersten Einzelarbeit (Brandenburg & Papenberg, 2022, S. 22-23) diskutiert.

Genauigkeit von NEST in Abhängigkeit zur Skalierung gemessener Variablen. Die Relevanz der Variablenskalierung für NEST liegt darin, dass die Variablenskalierung die Berechnung von Stichprobenkorrelationsmatrizen beeinflusst und somit die Genauigkeit von NEST unabhängig von Faktorenmodellen auf Populationsebene beeinträchtigen kann.

Im Speziellen wird im weiteren Verlauf eine simulationsbasierte Untersuchung und Modifikation von NEST für ordinale kategoriale Variablen[14] behandelt. Kategoriale Variablen sind in der empirischen psychologischen Forschung weit verbreitet; beispielsweise in Form binärcodierter Item-Lösungen oder in Form mehrstufiger Likert-Skalen. In Faktorenanalysen ergeben sich jedoch spezifische Probleme für kategoriale Variablen, unter anderem bei der Parameterschätzung (Flora & Curran, 2004), der Modellinterpretation (McDonald & Ahlawat, 1974; Olsson, 1979a) und der Bestimmung der Faktorenzahl (Cho et al., 2009; Cosemans et al., 2022; Garrido et al., 2013; Garrido et al. 2016; Goretzko & Bühner, 2022; Green et al., 2016; Lubbe, 2019; Timmerman & Lorenzo-Seva, 2011; Tran & Formann, 2009; Weng & Cheng, 2005, 2017; Yang & Xia, 2015).

Ein Ausgangspunkt für Probleme in Faktorenanalysen für kategoriale Variablen ist, dass die gängige Berechnung von Produkt-Moment-Korrelationen – im Nachfolgenden als *Pearson-Korrelationen* bezeichnet – für kategoriale Variablen andere statistische Eigenschaften als für kontinuierliche Variablen hat. Die erwartungstreue Berechnung von Pearson-Korrelationen für kontinuierliche Variablen ist nicht notwendigerweise erwartungstreu für kategoriale Variablen: Wenn zwei gemessene kategoriale Variablen als Operationalisierungen für zwei naturgemäß kontinuierliche Konstrukte verwendet werden, unterschätzt die Pearson-Korrelation der kategorialen Variablen den wahren Zusammenhang der kontinuierlichen Konstrukte (Garrido et al., 2013; Green et al., 2016; Lubbe, 2019).

Die ursprüngliche Implementierung von NEST, die Achim (2017) veröffentlichte, berechnet Stichprobenkorrelationsmatrizen in Form von Pearson-Korrelationen für alle untersuchten Datensätze, da die ursprüngliche Implementierung für Anwendungen mit kontinuierlichen Variablen entwickelt wurde. Wenn die ursprüngliche NEST-Implementierung auf Datensätze mit kategorialen Variablen angewendet wird, werden folglich ebenfalls Pearson-

[14] Die vorliegende Dissertation behandelt ausschließlich ordinale kategoriale Variablen und keine nominalskalierten Variablen. Daher werden ordinale kategoriale Variablen im Nachfolgenden vereinfachend als »kategoriale Variablen« bezeichnet, sind jedoch nicht zu verwechseln mit nominalskalierten Variablen.

Korrelationen für kategoriale Variablen berechnet. Die Unterschätzung des Zusammenhangs kategorialer Variablen durch Pearson-Korrelationen tritt demnach in der ursprünglichen NEST-Implementierung auf. Die simulierten Referenzdatensätze, die in NEST zur Simulation einer Stichprobenverteilung getesteter Eigenwerte benötigt werden, enthalten in der ursprünglichen Implementierung ausschließlich kontinuierliche Variablen. Daher werden für die simulierten Referenzdatensätzen ebenfalls Pearson-Korrelationen berechnet.

Da in den Simulationen 1a bis 2f ausschließlich Datensätze mit normalverteilten kontinuierlichen Variablen simuliert wurden, konnte die ursprüngliche NEST-Implementierung von Achim (2017) bedenkenlos übernommen werden. Inwiefern die positive Bewertung von NEST auch auf Datensätze mit kategorialen Variablen zutrifft, ist aufgrund der veränderten statistischen Eigenschaften der Korrelationsberechnung jedoch unklar und wurde daher in einer weiteren Simulation (Simulation 3) untersucht. Die vorliegende Dissertation adressiert damit eine Lücke in der bisherigen Validierung von NEST, die sich auf Anwendungen mit Pearson-Korrelationen für kontinuierliche Variablen beschränkt (Achim 2017; Brandenburg & Papenberg, 2022). In den folgenden Abschnitten werden die Implikationen unterschätzter Korrelationen für die ursprüngliche NEST-Implementierung zusammengefasst. Darauf aufbauend wird eine Modifikation von NEST mit einer angepassten Korrelationsberechnung für kategoriale Variablen erörtert. In Simulation 3 wurde die ursprüngliche NEST-Implementierung mit der modifizierten Implementierung verglichen.

Im Nachfolgenden werden Eigenwerte, die mit Faktoren korrespondieren, als *Signaleigenwerte* bezeichnet, da NEST die getestete Nullhypothese für diese Eigenwerte zurückweisen sollte. Komplementär dazu werden Eigenwerte, die nicht mit Faktoren korrespondieren, als *Rauscheigenwerte* bezeichnet. Die Unterschätzung von Korrelationen ist gleichbedeutend mit einer Unterschätzung geteilter Varianz und impliziert eine Unterschätzung der Kommunalitäten gemessener Variablen (siehe Gleichung 2), die durch Signaleigenwerte angegeben werden. Daher sind Signaleigenwerte bei Verwendung von Pearson-Korrelationen für kategoriale Variablen im Vergleich zu erwartungstreuen Stichprobenkorrelationen reduziert (Lubbe, 2019). Eine verringerte Signalstärke getesteter Eigenwerte legt nahe, dass NEST die optimale Faktorenzahl für kategoriale Variablen mit Pearson-Korrelationen häufiger unterschätzt als für kontinuierliche Variablen.

Präziser kann für NEST antizipiert werden, dass die optimale Faktorenzahl für kategoriale Variablen häufiger unterschätzt wird, je stärker die Verringerung von Signaleigenwerten ausfällt. Die Verringerung von Signaleigenwerten ist für binäre Variablen stärker als für Variablen mit mehr Antwortkategorien (Green et al., 2016). Außerdem ist die Verringerung von Signaleigenwerten stärker bei asymmetrischen Wahrscheinlichkeitsverteilungen der Antwortkategorien als bei symmetrischen Wahrscheinlichkeitsverteilungen – sowohl bei konstanter Asymmetrie in allen Variablen eines Datensatzes als auch bei variierender Asymmetrie (Lubbe, 2019)[15].

Neben möglichen Unterschätzungen der optimalen Faktorenzahl durch NEST mit Pearson-Korrelationen für kategoriale Variablen legen weitere theoretische Überlegungen außerdem mögliche Überschätzungen der optimalen Faktorenzahl nahe. Bei variierenden Wahrscheinlichkeitsverteilungen der Antwortkategorien zwischen den Variablen eines Datensatzes können Pearson-Korrelationen zwischen Variablenpaaren mit übereinstimmenden Wahrscheinlichkeitsverteilungen beobachtet werden (Green et al., 2016; Tran & Formann, 2009; Yang & Xia, 2015). In Datensätzen unter Faktorenmodellen äußern sich diese zusätzlichen Pearson-Korrelationen durch vergrößerte Rauscheigenwerte der Stichprobenkorrelationsmatrix (Garrido et al., 2013; Lim & Jahng, 2019). Da Kategoriewahrscheinlichkeiten in einem psychometrischen Kontext mit Item-Schwierigkeiten korrespondieren, werden Kategoriewahrscheinlichkeiten als Erklärung für beobachtete Korrelationen auch als *Difficulty Factors* (siehe Olsson, 1979a) bezeichnet. Eine Überschätzung der optimalen Faktorenzahl für kategoriale Variablen durch die ursprüngliche NEST-Implementierung wird insbesondere dadurch nahegelegt, dass die kontinuierliche Referenzvariablen die mögliche Anwesenheit von Difficulty Factors bei der Simulation der Stichprobenverteilung getesteter Eigenwerte nicht abbilden.

Die Verringerung von Signaleigenwerten und der mögliche Einfluss von Difficulty Factors implizieren die Vorhersage, dass die ursprüngliche NEST-Implementierung für kategoriale Variablen ungenauer als für kontinuierliche Variablen ist. Ob und in welchem Ausmaß diese Vorhersage anhand von Daten bestätigt werden kann, bedarf jedoch erneut umfangreicher Simulationen. In Simulation 3 wurden daher Datensätze mit kategorialen Variablen unter Faktorenmodellen simuliert, um die Lösungen der ursprünglichen NEST-

[15] Abbildung 1 der zweiten Einzelarbeit (Brandenburg, 2022, S. 12) veranschaulicht die Verringerung von Signaleigenwerten bei Pearson-Korrelationen für kategoriale Variablen als Funktion der Anzahl der Antwortkategorien und der Wahrscheinlichkeitsverteilungen der Antwortkategorien.

Implementierung mit Pearson-Korrelationen und kontinuierlichen Variablen in Referenzdatensätzen zu untersuchen, wenn die ursprüngliche Implementierung auf kategoriale Variablen angewendet wird. Basierend auf den Befunden, dass die Verringerung von Signaleigenwerten und die Anwesenheit von Difficulty Factors von der Anzahl (Green et al., 2016) und den Wahrscheinlichkeitsverteilungen (Lubbe, 2019) von Antwortkategorien abhängen, wurden in Simulation 3 sowohl die Anzahl der Antwortkategorien als auch deren Wahrscheinlichkeitsverteilungen gezielt manipuliert (siehe unten).

Polychorische Korrelationen

Als Alternative zur ursprünglichen NEST-Implementierung wurde für die vorliegende Dissertation eine modifizierte NEST-Implementierung entwickelt, die *polychorische Korrelationen* für kategoriale Variablen berechnet. Polychorische Korrelationen berechnen die Korrelation angenommener latenter kontinuierlicher Repräsentationen der gemessenen kategorialen Variablen (für technische Details der Berechnung, siehe Lubbe, 2019; Olsson, 1979b). Im Gegensatz zu Pearson-Korrelationen für kategoriale Variablen sind polychorische Korrelationen erwartungstreu unter einer Normalverteilungsannahme bezüglich der latenten Repräsentationen (Lubbe, 2019; Muthén, 1978). Außerdem reflektieren polychorische Korrelationen keine Difficulty Factors (Garrido et al., 2013; Tran & Formann, 2009; Yang & Xia, 2015). Die Motivation, eine modifizierte NEST-Implementierung mit polychorischen Korrelationen zur Berechnung von Stichprobenkorrelationsmatrizen zu entwickeln, ergibt sich demnach daraus, dass polychorische Korrelationen die skizzierten Probleme der ursprünglichen Implementierung durch unterschätzte Korrelationen und Difficulty Factors umgehen.

Ein wichtiges Merkmal der modifizierten NEST-Implementierung mit polychorischen Korrelationen ist, dass die Referenzdatensätze zur Simulation der Stichprobenverteilung getesteter Eigenwerte ebenfalls kategoriale Variablen enthalten. Polychorische Korrelationen werden sowohl für die gemessenen kategorialen Variablen als auch für die simulierten kategorialen Variablen der Referenzdatensätze berechnet[16]. Die Berechnung polychorischer Korrelationen für Referenzdatensätze ist notwendig, um übereinstimmende statistische Eigenschaften

[16] Technische Details zum Umgang mit nichtdefiniten Korrelationsmatrizen bei polychorischen Korrelationen für große Mengen an Referenzdatensätzen (Green et al., 2016; Roznowski et al., 1991; Timmerman & Lorenzo-Seva, 2011; Tran & Formann, 2009) werden in der zweiten Einzelarbeit (Brandenburg, 2022, S. 42-43) erläutert.

der berechneten Stichprobenkorrelationen für die gemessenen Variablen und die Referenzdatensätze zu erreichen. Auf diese Weise wird in der modifizierten NEST-Implementierung eine adäquate Simulation der Stichprobenverteilung getesteter Eigenwerte unter der Nullhypothese in NEST sichergestellt.

Trotz der genannten Vorteile polychorischer Korrelationen gegenüber Pearson-Korrelationen erfordert die modifizierte NEST-Implementierung eine simulationsbasierte Validierung inklusive eines Vergleichs mit der ursprünglichen NEST-Implementierung. Ein Nachteil polychorischer Korrelationen gegenüber Pearson-Korrelationen ist, dass polychorische Korrelationen einen größeren Standardfehler haben (Garrido et al., 2013; Garrido et al., 2016; Timmerman & Lorenzo-Seva, 2011; Tran & Formann, 2009; Weng & Cheng, 2017). Für Datensätze unter Faktorenmodellen impliziert ein größerer Standardfehler berechneter Korrelationen eine größere Dispersion der Signal- und Rauscheigenwerte von Stichprobenkorrelationsmatrizen (Lubbe, 2019). Die größere Dispersion von Eigenwerten führt dazu, dass die Signalstärke einzelner Signaleigenwerte in NEST geringer ist als bei der Berechnung von Pearson-Korrelationen für kontinuierliche Variablen unter äquivalenten Faktorenmodellen. Lubbe (2019) zeigte außerdem, dass die Dispersion der Eigenwerte bei asymmetrischen Wahrscheinlichkeitsverteilungen der Antwortkategorien größer als bei symmetrischen Wahrscheinlichkeitsverteilungen war.

Die Abhängigkeit der Dispersion der Eigenwerte von den Wahrscheinlichkeitsverteilungen gemessener Variablen wird in der modifizierten NEST-Implementierung mit polychorischen Korrelationen berücksichtigt. Die simulierten Variablen der Referenzdatensätze reproduzieren die Kategoriewahrscheinlichkeiten der gemessenen Variablen, um die Dispersion der simulierten Stichprobenverteilung getesteter Eigenwerte unter der Nullhypothese dem Standardfehler der berechneten polychorischen Korrelationen für die gemessenen Variablen anzupassen (parallel zu Lubbe, 2019)[17].

Die verringerte Signalstärke von Signaleigenwerten aufgrund erhöhter Dispersion von Eigenwerten bei polychorischen Korrelationen legt nahe, dass die modifizierte NEST-Implementierung die optimale Faktorenzahl für kategoriale

[17] Das Vorgehen in der modifizierten NEST-Implementierung, polychorische Korrelationen zu verwenden und beobachtete Kategoriewahrscheinlichkeiten in simulierten Referenzdatensätzen zu reproduzieren, orientiert sich an der PA-Implementierung für kategoriale Variablen von Lubbe (2019). Es wurde bisher nicht untersucht, inwiefern NEST für kategoriale Variablen durch ein ähnliches Vorgehen profitiert.

Variablen trotz erwartungstreuer polychorischer Korrelationen häufiger unterschätzt als die ursprüngliche NEST-Implementierung für kontinuierliche Variablen. Außerdem kann antizipiert werden, dass Unterschätzungen mit asymmetrischen Wahrscheinlichkeitsverteilungen häufiger auftreten als mit symmetrischen Wahrscheinlichkeitsverteilungen. In Simulation 3 untersucht wurde untersucht, in welchem Umfang sich diese Überlegungen bestätigen lassen. Außerdem wurde in Simulation 3 untersucht, ob die modifizierte NEST-Implementierung für kategoriale Variablen trotz der genannten Probleme mit polychorischen Korrelationen insgesamt genauer für kategoriale Variablen als die ursprüngliche NEST-Implementierung ist.

Simulation 3

Simulation 3 wurde an die Simulation von Yang und Xia (2015) angelehnt, die zur Untersuchung von Methoden zur Bestimmung der Faktorenzahl für kategoriale Variablen ausgearbeitet werden war. Die Simulation von Yang und Xia war für die vorliegenden Fragestellungen geeignet, da sie Manipulationen der Anzahl der Antwortkategorien pro Variable und deren Wahrscheinlichkeitsverteilungen beschreibt. Wie bereits die Simulationen 1a bis 2f beinhaltete Simulation 3 vollständig-gekreuzte Manipulationen. Die Manipulationen in Simulation 3 dienten sowohl der gezielten Untersuchung der problematischen Datenmerkmale (Anzahl der Antwortkategorien, Wahrscheinlichkeitsverteilungen der Antwortkategorien) als auch der Simulation heterogener Anwendungsfälle. Konkret wurden in Simulation 3 die Manipulationen der Anzahl der Antwortkategorien (zwei Antwortkategorien, vier Antwortkategorien) und der Wahrscheinlichkeitsverteilungen der Antwortkategorien (symmetrisch, konstant-asymmetrisch, variierend-asymmetrisch) von Yang und Xia übernommen. Die spezifizierte Anzahl der Antwortkategorien galt dabei für alle Variablen des jeweils simulierten Datensatzes.

Die Simulation kategorialer Variablen erfolgte in Simulation 3 so, dass zunächst multivariat-normalverteilte Datensätze unter den spezifizierten Faktorenmodellen simuliert wurden. Anschließend wurden die normalverteilten Variablen nach der Methode von Yang und Xia (2015) anhand vordefinierter Schwellenwerte in kategoriale Variablen transformiert, um die spezifizierten diskreten Wahrscheinlichkeitsverteilungen der Antwortkategorien auf Populationsebene zu erreichen (für eine technische Beschreibung der Transformation und für die exakten Kategoriewahrscheinlichkeiten, siehe Brandenburg, 2022, S. 11; Yang & Xia, 2015). Orthogonal zur Anzahl der Antwortkategorien und deren

Wahrscheinlichkeitsverteilungen wurde unter anderem die Stichprobengröße der simulierten Datensätze manipuliert ($N = 100$, $N = 500$; für das vollständige Design mit allen manipulierten Datenmerkmalen und deren Faktorstufen, siehe Brandenburg, 2022, S. 15).

Die ursprüngliche NEST-Implementierung und die modifizierten NEST-Implementierung wurden auf jeden simulierten Datensatz angewendet. Im Nachfolgenden werden die ursprüngliche Implementierung als $NEST_{Pearson}$ und die modifizierte Implementierung als $NEST_{poly}$ bezeichnet. Als Referenzmethode für die NEST-Implementierungen wurde in Simulation 3 erneut PA verwendet. Die hierbei verwendete PA-Implementierung wurde von Lubbe (2019) veröffentlicht und parallel dazu in einer Simulation als optimale PA-Implementierung für kategoriale Variablen befunden. Diese PA-Implementierung basiert ebenso wie $NEST_{poly}$ auf polychorischen Korrelationen und wird daher als PA_{poly} bezeichnet. Die Bewertung aller Methodenlösungen nach den vier Kategorien der Simulationen 1a bis 2f (akkurat, unterschätzt, überschätzt, undefiniert) wurde in Simulation 3 übernommen.

Zusätzlich wurde Simulation 3 um eine supplementäre Simulation ergänzt, in der Datensätze mit kontinuierlichen Variablen unter den gleichen Manipulationen wie in Simulation 3 simuliert wurden – ausgenommen der Manipulationen der Antwortkategorien und deren Wahrscheinlichkeitsverteilungen. In der supplementären Simulation wurde $NEST_{Pearson}$ auf alle simulierten Datensätze angewendet, um für die Bedingungen von Simulation 3 ein Vergleichsniveau der Genauigkeit von NEST für kontinuierliche Variablen herzustellen.

Ergebnisse der Simulation 3

Die Ergebnisse der Simulation 3 zeigen, dass relative Präferenzen für $NEST_{Pearson}$ oder $NEST_{poly}$ für kategoriale Variablen von der Anzahl der Antwortkategorien, den Wahrscheinlichkeitsverteilungen und der Stichprobengröße abhingen. Daher werden die Ergebnisse getrennt nach diesen Datenmerkmalen berichtet.

Die Ergebnisse der untersuchten Methoden für Datensätze mit binären Variablen (zwei Antwortkategorien) sind in Tabelle 3 zusammengefasst. Alle Methoden waren für binäre Variablen am genauesten mit Stichprobengrößen von $N = 500$ und symmetrischen Wahrscheinlichkeitsverteilungen. Die Ergebnisse weisen auf einen Genauigkeitsvorteil von $NEST_{poly}$ für binäre Variablen gegenüber

$NEST_{Pearson}$ und PA_{poly} hin, der jedoch auf Datensätze mit $N = 500$ beschränkt war. Mit $N = 500$ war $NEST_{poly}$ bei allen Wahrscheinlichkeitsverteilungen genauer als $NEST_{Pearson}$ und PA_{poly} und insbesondere im Vergleich zu $NEST_{Pearson}$ robuster gegen manipulierte Wahrscheinlichkeitsverteilungen. Mit $N = 100$ war PA_{poly} die genaueste Methode für binäre Variablen. Insgesamt konnte mit $N = 100$ jedoch keine Methode eine Genauigkeit von 50.0 % oder mehr erzielen.

Tabelle 3

Lösungen für binäre Variablen in Simulation 3 als Funktion der Stichprobengröße und der Wahrscheinlichkeitsverteilungen der Antwortkategorien

Methode	Lösung	$N = 100$			$N = 500$		
		symmetrisch	konstant-asymmetrisch	variierend-asymmetrisch	symmetrisch	konstant-asymmetrisch	variierend-asymmetrisch
$NEST_{Pearson}$	überschätzt	2.8	16.1	11.5	3.7	17.1	44.1
	akkurat	42.6	32.5	18.4	75.9	58.9	21.1
	unterschätzt	54.6	51.4	70.1	20.4	24.0	34.8
	undefiniert	0	0	0	0	0	0
$NEST_{poly}$	überschätzt	1.0	0.5	0.6	2.2	1.6	3.0
	akkurat	39.7	26.4	18.8	77.1	64.7	59.9
	unterschätzt	59.3	72.3	80.7	20.6	33.7	37.1
	undefiniert	0	0.8	0	0	0	0
PA_{poly}	überschätzt	9.4	12.1	9.4	2.2	7.7	8.9
	akkurat	41.9	36.4	30.0	63.5	52.1	49.7
	unterschätzt	48.6	51.5	60.6	34.3	40.2	41.4
	undefiniert	0	0	0	0	0	0

Bemerkung: Die Werte geben den prozentualen Anteil der jeweiligen Lösungskategorie an den Lösungen der entsprechenden Methode für binäre Variablen (insgesamt 9,600 simulierte Datensätze) an.

Die Ergebnisse für Variablen mit vier Antwortkategorien sind in Tabelle 4 zusammengefasst. Für Variablen mit vier Antwortkategorien waren die Methoden – mit Ausnahme von $NEST_{poly}$ – erneut am genauesten mit $N = 500$ und symmetrischen Wahrscheinlichkeitsverteilungen. $NEST_{Pearson}$ und PA_{poly} waren für Variablen mit vier Antwortkategorien insgesamt genauer als für binäre Variablen. $NEST_{poly}$ hingegeben war für Variablen mit vier Antwortkategorien ungenauer als für binäre Variablen und ungenauer als $NEST_{Pearson}$ oder PA_{poly}. Für Variablen mit vier Antwortkategorien war $NEST_{Pearson}$ insgesamt die genaueste Methode; die einzige Ausnahme waren Datensätze mit $N = 500$ und variierend-asymmetrischen Wahrscheinlichkeitsverteilungen. PA_{poly} zeigte für Variablen mit vier

Antwortkategorien die größte Robustheit gegen manipulierte Wahrscheinlichkeitsverteilungen.

Tabelle 4

Lösungen für Variablen mit vier Antwortkategorien in Simulation 3 als Funktion der Stichprobengröße und der Wahrscheinlichkeitsverteilungen der Antwortkategorien

Methode	Lösung	$N = 100$			$N = 500$		
		symmetrisch	konstant-asymmetrisch	variierend-asymmetrisch	symmetrisch	konstant-asymmetrisch	variierend-asymmetrisch
$NEST_{Pearson}$	überschätzt	1.3	3.1	5.1	1.7	4.1	37.0
	akkurat	59.1	53.7	46.7	85.1	80.7	46.2
	unterschätzt	39.7	43.2	48.2	13.2	15.2	16.8
	undefiniert	0	0	0	0	0	0
$NEST_{poly}$	überschätzt	0	0	0.1	0	0	0
	akkurat	25.6	30.0	29.4	45.1	70.4	45.2
	unterschätzt	74.4	70.0	70.5	54.9	29.6	54.8
	undefiniert	0	0	0	0	0	0
PA_{poly}	überschätzt	5.8	7.0	8.3	0	0.4	0.2
	akkurat	48.9	46.3	46.2	70.4	68.1	68.6
	unterschätzt	45.3	46.7	45.5	29.6	31.5	31.2
	undefiniert	0	0	0	0	0	0

Bemerkung: Die Werte geben den prozentualen Anteil der jeweiligen Lösungskategorie an den Lösungen der entsprechenden Methode für Variablen mit vier Antwortkategorien (insgesamt 9,600 simulierte Datensätze) an.

In der supplementären Simulation wurden die Lösungen von $NEST_{Pearson}$ für kontinuierliche Variablen in vergleichbaren Bedingungen wie in Simulation 3 ermittelt. $NEST_{Pearson}$ war für kontinuierliche Variablen mit $N = 500$ (88.8 % akkurat; 10.0 % unterschätzt, 1.2 % überschätzt) genauer als mit $N = 100$ (63.1 % akkurat, 35.0 % unterschätzt, 1.9 % überschätzt). Der Vergleich der Ergebnisse für kontinuierliche und kategoriale Variablen zeigt, dass beide NEST-Implementierungen für kategoriale Variablen stets ungenauer waren als $NEST_{Pearson}$ für kontinuierliche Variablen mit jeweils gleicher Stichprobengröße.

Die verringerte Genauigkeit für kategoriale Variablen äußerte sich in $NEST_{Pearson}$ und $NEST_{poly}$ darin, dass beide NEST-Implementierungen die optimale Faktorenzahl häufiger unterschätzten als $NEST_{Pearson}$ für kontinuierliche Variablen mit jeweils gleicher Stichprobengröße. Häufigere Unterschätzungen der optimalen Faktorenzahl für kategoriale Variablen sind konsistent mit den theoretischen Überlegungen, dass die Signalstärke getesteter Signaleigenwerte mit Pearson-Korrelationen und polychorischen Korrelationen für kategoriale Variablen verringert ist. In Simulation 3

unterschätzte $NEST_{poly}$ die optimale Faktorenzahl häufiger als $NEST_{Pearson}$ (siehe Tabellen 3 und 4). Häufigere Unterschätzungen durch $NEST_{poly}$ deuten darauf hin, dass die Reduktion der Signalstärke für Signaleigenwerte aufgrund des Standardfehlers polychorischer Korrelationen schwerwiegender war als die Reduktion der Signalstärke aufgrund unterschätzter Pearson-Korrelationen.

Neben vermehrter Unterschätzungen der optimalen Faktorenzahl zeigte $NEST_{Pearson}$ für kategoriale Variablen mit asymmetrischen Wahrscheinlichkeitsverteilungen zudem häufigere Überschätzungen im Vergleich zu kontinuierlichen Variablen (siehe Tabellen 3 und 4). Überschätzungen durch $NEST_{Pearson}$ nahmen mit zunehmender Stichprobengröße zu. Auffällige Überschätzungen durch $NEST_{Pearson}$ speziell für Datensätze mit asymmetrisch-verteilten kategorialen Variablen sind konsistent mit der Überlegung, dass $NEST_{Pearson}$ anfällig für die Akzeptanz von Difficulty Factors ist. Die im Vergleich zu $NEST_{Pearson}$ selteneren Überschätzungen durch $NEST_{poly}$ sind zudem konsistent Überlegung, dass Überschätzungen aufgrund von Difficulty Factors einseitig $NEST_{Pearson}$ betreffen.

Diskussion der Simulation 3

Simulation 3 wurde durchgeführt, um die Bestimmung der Faktorenzahl für kategoriale Variablen durch NEST zu untersuchen und dabei Vor- und Nachteile von Pearson-Korrelationen und polychorischen Korrelationen aufzudecken. Dazu wurden kategoriale Variablen unter Faktorenmodellen simuliert und die Lösungen von zwei NEST-Implementierungen – $NEST_{Pearson}$ und $NEST_{poly}$ – verglichen. Grundsätzlich zeigten sowohl $NEST_{Pearson}$ als auch $NEST_{poly}$ für kategoriale Variablen häufigere Unterschätzungen der optimalen Faktorenzahl als $NEST_{Pearson}$ für kontinuierliche Variablen. $NEST_{Pearson}$ zeigte gleichzeitig häufigere Überschätzungen bei asymmetrischen Wahrscheinlichkeitsverteilungen der Antwortkategorien. Damit zeigen die Ergebnisse der Simulation 3 zusätzliche Fehlerquellen für NEST, die über die Erkenntnisse der Simulationen 1a bis 2f hinausgehen, da NEST in den Simulationen 1a bis 2f ausschließlich mit Pearson-Korrelationen für kontinuierliche Variablen untersucht wurde.

Die Ergebnisse der Simulation 3 sind uneindeutig im Hinblick auf die Fragestellung, welche Methode insgesamt für kategoriale Variablen präferiert werden kann. Die Genauigkeit von $NEST_{Pearson}$ und $NEST_{poly}$ war in Simulation 3 abhängig von verschiedenen Datenmerkmalen. Keine Implementierung kann vor

dem Hintergrund der Ergebnisse als dominant für kategoriale Variablen bezeichnet werden. Für binäre Variablen mit $N = 500$ war $NEST_{poly}$ trotz häufigerer Unterschätzungen der optimalen Faktorenzahl als $NEST_{Pearson}$ insgesamt die genauere NEST-Implementierung. Da $NEST_{poly}$ für binäre Variablen mit $N = 500$ zusätzlich genauer als PA_{poly} war, wird ebenfalls die Präferenz von $NEST_{poly}$ gegenüber PA_{poly} für binäre Variablen mit $N = 500$ unterstützt. Der Genauigkeitsvorteil von $NEST_{poly}$ gegenüber PA_{poly} ergänzt daher die Erkenntnisse von Lubbe (2019), die PA_{poly} als präferierte Methode zur Bestimmung der Faktorenzahl für kategoriale Variablen auszeichneten. Ein Mehrwert von $NEST_{poly}$ gegenüber $NEST_{Pearson}$ oder PA_{poly} zeigte sich jedoch nicht für binäre Variablen mit $N = 100$ oder Variablen mit vier Antwortkategorien.

Für Variablen mit vier Antwortkategorien war $NEST_{Pearson}$ genauer als $NEST_{poly}$ und PA_{poly}, jedoch limitiert auf Datensätze in denen Wahrscheinlichkeitsverteilungen zwischen Variablen nicht variierten. Übersetzt in Item-Schwierigkeiten bedeutet dieses Ergebnis, dass $NEST_{Pearson}$ die präferierte Methode für kategoriale Variablen mit vier Antwortkategorien war, wenn Variablen homogene Item-Schwierigkeiten aufwiesen. Bei heterogenen Item-Schwierigkeiten war $NEST_{Pearson}$ anfällig für die Akzeptanz von Difficulty Factors.

Die mangelnde Eignung von $NEST_{Pearson}$ für Variablen mit heterogenen Item-Schwierigkeiten wird zusätzlich dadurch untermauert, dass Überschätzungen durch $NEST_{Pearson}$ in Simulation 3 bei heterogenen Item-Schwierigkeiten am häufigsten mit $N = 500$ anstatt mit $N = 100$ auftraten. Mit zunehmender Stichprobengröße sinkt der Standardfehler von Pearson-Korrelationen und somit auch die Dispersion der Eigenwerte von Stichprobenkorrelationsmatrizen (siehe oben). Zunehmende Anfälligkeit für Difficulty Factors durch $NEST_{Pearson}$ bei geringerer Dispersion der getesteten Eigenwerte zeigt, dass die simulierte Stichprobenverteilung getesteter Eigenwerte in $NEST_{Pearson}$ den Einfluss von Difficulty Factors nicht abbildet[18]. Difficulty Factors sind als Problem in Faktorenanalysen bereits bekannt (siehe Garrido et al., 2013; McDonald & Ahlawat, 1974; Olsson, 1979a), aber erst die Ergebnisse der Simulation 3 zeigen, dass Difficulty Factors die Genauigkeit von

[18] In der zweiten Einzelarbeit (Brandenburg, 2022, S. 26) wird anhand einer zusätzlichen Simulation bestätigt, dass es sich bei der Zunahme von Überschätzungen der optimalen Faktorenzahl durch $NEST_{Pearson}$ mit zunehmender Stichprobengröße um eine Verfehlung einer adäquaten Stichprobenverteilung getesteter Eigenwerte handelte. Durch die zusätzliche Simulation kann ausgeschlossen werden, dass es sich bei den vermehrten Überschätzungen bei $N = 500$ um ein simples Artefakt erhöhter Exposition zu Rauscheigenwerten aufgrund seltenerer Unterschätzungen handelte.

$NEST_{Pearson}$ reduzieren, da in $NEST_{Pearson}$ bei heterogenen Item-Schwierigkeiten keine geeignete Stichprobenverteilung getesteter Rauscheigenwerte simuliert wird.

PA_{poly} war robuster als beide NEST-Implementierungen gegen manipulierte Wahrscheinlichkeitsverteilungen und am genauesten für Variablen mit vier Antwortkategorien und variierend-asymmetrischen Wahrscheinlichkeitsverteilungen (siehe Tabellen 3 und 4). Eine Präferenz für NEST gegenüber PA für nicht-binäre kategoriale Variablen mit heterogenen Item-Schwierigkeiten kann damit nicht durch die vorliegenden Ergebnisse gerechtfertigt werden.

Für Variablen mit vier Antwortkategorien und homogenen Item-Schwierigkeiten (symmetrische und konstant-asymmetrische Wahrscheinlichkeitsverteilungen) näherte sich die Genauigkeit von $NEST_{Pearson}$ den Ergebnissen für kontinuierliche Variablen an. Die Annäherung an das Niveau für kontinuierliche Variablen bei Verwendung von Pearson-Korrelationen ist konsistent damit, dass Pearson-Korrelationen den wahren Zusammenhang kategorialer Variablen mit zunehmender Anzahl an Antwortkategorien weniger stark unterschätzen (Green et al., 2016). Da sich $NEST_{poly}$ für Variablen mit vier Antwortkategorien im Vergleich zu binären Variablen nicht verbesserte, liegt nahe, dass $NEST_{Pearson}$ auch für kategoriale Variablen mit mehr als vier Antwortkategorien die präferierte NEST-Implementierung ist.

Die reduzierte Genauigkeit aller untersuchten Methoden mit $N = 100$ relativ zu $N = 500$ untermauert, dass die Bestimmung der Faktorenzahl große Stichproben erfordert und Datenerhebungen zu diesem Zweck entsprechend geplant werden sollten. Die vorliegenden Ergebnisse stimmen darin mit weiteren Simulationsstudien überein, in denen erhöhte Genauigkeit bei der Bestimmung der Faktorenzahl durch verschiedene Methoden mit zunehmender Stichprobengröße berichtetet wurde (Auerswald & Moshagen, 2019; Braeken & van Assen, 2017; Garrido et al., 2013; Green et al., 2012; Keith et al., 2016; Li et al., 2020; Lim & Jahng, 2019; Ruscio & Roche, 2012). Die niedrige Genauigkeit bei $N = 100$ stimmt darüber hinaus mit aktuellen Empfehlungen für explorative Faktorenanalysen mit einer Mindeststichprobengröße von $N = 400$ (Goretzko et al., 2021) überein.

Die uneindeutigen Ergebnisse bezüglich der präferierten NEST-Implementierung für kategoriale Variablen sollten in zukünftigen Simulationsstudien zum Faktorenzahlproblem mit Bezug zu NEST berücksichtigt werden. Lim und Jahng (2019) untersuchten eigenwertbasierte Methoden zur

Bestimmung der Faktorenzahl für kategoriale Variablen, die Eigenwerte ähnlich wie NEST testen, aber nutzten dabei ausschließlich polychorische Korrelationen zur Berechnung von Stichprobenkorrelationsmatrizen. Die Ergebnisse der Simulation 3 legen jedoch nahe, dass polychorische Korrelationen für kategoriale Variablen nicht notwendigerweise einen Mehrwert zur Bestimmung der Faktorenzahl durch eigenwertbasierte Methoden wie NEST bieten. Vergleichende Studien unter Berücksichtigung von NEST oder ähnlichen Methoden sollten den vorliegenden Ergebnissen zufolge unterschiedliche Implementierungen mit Pearson- und polychorischen Korrelationen berücksichtigen (siehe Cosemans et al., 2022).

Eine Limitation von Simulation 3 ist, dass Datensätze ausschließlich mit derselben Anzahl an Antwortkategorien pro Variable simuliert wurden. Durch Simulation 3 wird daher nicht dazu beigetragen, optimale Vorgehensweisen zur Bestimmung der Faktorenzahl für Datensätze mit unterschiedlich skalierten Variablen zu identifizieren – beispielsweise für Datensätze, die sowohl kategoriale als auch kontinuierliche Variablen enthalten. Es existieren Softwarelösungen zur Korrelationsberechnung in gemischt-skalierten Datensätzen, die unterschiedliche Korrelationsmaße in Korrelationsmatrizen abhängig von den Skalierungen des jeweiligen Variablenpaares vorsehen (Beispiel: Revelle, 2022). Inwiefern gemischte Korrelationsmatrizen die jeweiligen Vor- und Nachteile unterschiedlicher Korrelationsberechnungen bei der Bestimmung der Faktorenzahl kombinieren, bedarf weiterer Forschung. Ebenso bedarf es weiterer Forschung zur optimalen Variablenskalierung und Korrelationsberechnung für Referenzdatensätze in NEST für gemischt-skalierte Datensätze.

Eine weitere Limitation betrifft die Normalverteilungsannahme bezüglich latenter kontinuierlicher Repräsentationen kategorialer Variablen in der Berechnung polychorischer Korrelationen (siehe Olsson, 1979b). Die Normalverteilungsannahme wurde Simulation 3 nicht verletzt aufgrund der Transformation normalverteilter Variablen zur Simulation kategorialer Variablen (siehe Yang & Xia, 2015). Die gleiche Normalverteilungsannahme wurde auch in den Referenzdatensätzen von $NEST_{poly}$ nicht verletzt. Jin und Yang-Wallentin (2017) zeigten in Simulationen, dass polychorische Korrelationen für kategoriale Variablen die wahren Zusammenhänge kontinuierlicher Konstrukte unterschätzen, wenn die kontinuierlichen Konstrukte die Normalverteilungsannahme bei der Berechnung polychorischer Korrelationen verletzen. Vor diesem Hintergrund bedarf es weiterer Forschung, ob $NEST_{poly}$ weiterhin für binäre Variablen präferiert werden kann, wenn die

Normalverteilungsannahme bei der Berechnung polychorischer Korrelationen verletzt ist. Eine potentielle Ursache für ungenaue Lösungen von $NEST_{poly}$ ist, dass simulierte Referenzdatensätze möglicherweise keine geeignete Stichprobenverteilung getesteter Eigenwerte abbilden, wenn getestete Eigenwerte aufgrund verletzter Normalverteilungsannahme reduziert sind und simulierte Referenzdatensätze diese Verletzung nicht reproduzieren.

Zusammengefasst zeigen die vorliegenden Ergebnisse, dass kategoriale Variablen die Genauigkeit der ursprünglichen NEST-Implementierung nachweislich reduzierten aufgrund unterschätzter Pearson-Korrelationen und missachteter Difficulty Factors. Bei Verwendung polychorischer Korrelationen und kategorialer Variablen in Referenzdatensätzen reduzierten stattdessen instabile Korrelationsberechnungen die Genauigkeit der modifizierten NEST-Implementierung. Der Vergleich der NEST-Implementierungen zeigt, dass Datenmerkmale die Abwägung der Vor- und Nachteile der Implementierungen bestimmen. Die Verwendung polychorischer Korrelationen anstelle von Pearson-Korrelationen resultierte in der modifizierten NEST-Implementierung in höherer Genauigkeit für binäre Variablen, jedoch limitiert auf große Stichproben (hier $N = 500$). Dagegen resultierte die Verwendung von Pearson-Korrelationen in der ursprünglichen NEST-Implementierung in höherer Genauigkeit für Variablen mit vier Antwortkategorien, jedoch limitiert auf homogene Item-Schwierigkeiten. Ein Genauigkeitsvorteil von NEST gegenüber PA konnte für Variablen mit vier Antwortkategorien und heterogenen Item-Schwierigkeiten nicht nachgewiesen werden.

Generelle Diskussion

Die optimale Lösung des Faktorenzahlproblems bei explorativen Faktorenanalysen soll den Anforderungen genügen, die Faktorenzahl für ein Faktorenmodell beobachteter Korrelationen einerseits passend und andererseits sparsam zu bestimmen (Auerswald & Moshagen, 2019; Cosemans et al., 2022; Fava & Velicer, 1992; Revelle & Rocklin, 1979; Zwick & Velicer, 1986). Die aktuelle Forschung zum Faktorenzahlproblem ist reich an Vorschlägen für Methoden zur Bestimmung der nach diesen Ansprüchen optimalen Faktorenzahl (Achim, 2017; Braeken & van Assen, 2017; Golino & Epskamp, 2017; Goretzko & Bühner, 2020; Green et al., 2012; Lorenzo-Seva et al., 2011; Ruscio & Roche, 2012). Die Validierung neuer Methoden ist

jedoch limitiert in den Bedingungen, in denen Methoden bisher untersucht worden sind, und in den Vergleichen mit weiteren Methoden.

Die Forschungsfragen der vorliegenden Dissertation sind diesen Limitationen gewidmet. Die vorliegenden Ergebnisse ergänzen die Validierung der untersuchten Methoden, indem Abhängigkeiten ihrer Lösungen von Datenmerkmalen mittels Simulationen gefunden wurden, die über zuvor betrachtete Bedingungen hinausgingen. In den vorliegenden Simulationen wurden außerdem verschiedene Methoden verglichen. Die Vergleiche der Methoden zeigt relevante Genauigkeitsunterschiede, durch die wiederum Präferenzen für einzelne Methoden begründet werden.

In den Simulationen 1a bis 2f wurden die Methoden NEST (Achim, 2017) und EGA (Golino & Epskamp, 2017) erstmals miteinander verglichen und zudem erstmals auf Robustheit gegen substantielle Kreuzladungen untersucht. NEST war bei der Bestimmung der Faktorenzahl in Anwesenheit substantieller Kreuzladungen bisweilen erheblich genauer als EGA, deren Genauigkeit bei kontrollierter Manipulation der Kreuzladungen stärker abnahm (Simulationen 1a und 1b) und mit Circumplexmodellen nicht vereinbar war (Simulationen 2a bis 2f). Diese Ergebnisse deuten darauf hin, dass EGA mehr als NEST die implizite Annahme voraussetzt, dass gemessene Variablen nicht durch substantielle Kreuzladungen determiniert werden. Diese Schlussfolgerung ist insbesondere ein relevanter Beitrag zur Validierung von NEST und EGA, da die impliziten Annahmen zu Kreuzladungen weder für NEST noch für EGA in vorherigen Arbeiten (Achim, 2017; Golino & Epskamp, 2017; Golino et al., 2020) behandelt worden sind. Im Vergleich zwischen NEST und EGA als konkurrierende Methoden unterstützen die Ergebnisse der Simulationen 1a bis 2f die Anwendung von NEST in explorativen Faktorenanalysen, da die impliziten Annahmen zu Kreuzladungen durch EGA restriktiver sind als die inhärenten Annahmen explorativer Faktorenanalysen (siehe Achim, 2020).

Anknüpfend an die positive Bewertung von NEST in den Simulationen 1a bis 2f wurde die Untersuchung von NEST um Bedingungen mit kategorialen Variablen erweitert. Die Ergebnisse der Simulation 3 zeigen, dass eine Präferenz der ursprünglichen NEST-Implementierung gegenüber der getesteten PA-Implementierung (Lubbe, 2019) für Likert-Skalen-artige Variablen weiterhin gerechtfertigt ist, sofern die gemessenen Variablen homogene Item-Schwierigkeiten aufweisen. Die Ergebnisse zeigen ein bisher unerkanntes Genauigkeitsdefizit der ursprünglichen NEST-Implementierung bei heterogenen Item-Schwierigkeiten.

Außerdem zeigen die Ergebnisse der Simulation 3 eine Verbesserung der Genauigkeit für binäre Variablen durch eine modifizierte NEST-Implementierung mit polychorischen Korrelationen und kategorialen Variablen in Referenzdatensätzen. Der Genauigkeitsvorteil der modifizierten NEST-Implementierung gegenüber der ursprünglichen Implementierung für binäre Variablen war jedoch auf große Stichproben ($N = 500$) limitiert.

In anderen Kontexten wurde der Gedanke diskutiert, Simulationsstudien zur Validierung quantitativer Methoden kritisch zu überprüfen, um vermeintlich optimistische Schlussfolgerung zu neuen Methoden anhand zusätzlicher Problemstellungen kontinuierlich neu zu bewerten (Boulesteix et al., 2017; Nießl et al., 2017). Die vorliegende Dissertation ist im Einklang mit der Forderung nach der Überprüfung veröffentlichter Simulationsstudien und enthält dahingehend Neubewertungen, dass die vorliegenden Ergebnisse die Anwendbarkeit von EGA und NEST strenger eingrenzen als vorherige Simulationsstudien (Achim 2017; Golino et al., 2020).

Die Neubewertungen auf Grundlage der vorliegenden Ergebnisse eröffnen Fragestellungen zur Untersuchung weiterer Methoden zur Bestimmung der Faktorenzahl (Beispiele: Braeken & van Assen, 2017, Goretzko & Bühner, 2020; Lorenzo-Seva et al., 2011; Ruscio & Roche, 2012). Ein Ansatz für weitere Forschung sind Vergleiche weiterer Methoden unter Bedingungen, die Genauigkeitsdefizite für Methoden nahelegen (siehe Kreuzladungen in den Simulationen 1a bis 2f und kategoriale Variablen in Simulation 3), um nach relevanten Genauigkeitsunterschieden zwischen Methoden zu suchen. Beispielsweise wird durch die Simulationen 1a bis 2f die Frage aufgeworfen, wie sich weitere Methoden in Anwesenheit substantieller Kreuzladungen vergleichen und ob sie – ähnlich wie NEST und EGA – unterschiedlich restriktive Annahmen zu Kreuzladungen voraussetzen. Darüber hinaus legt das uneindeutige Ergebnismuster der Simulation 3 bezüglich der optimalen Berechnung von Korrelationen für kategoriale Variablen nahe, die Verwendung von Pearson- und polychorischer Korrelationen für weitere Methoden gegenüberzustellen, die die Berechnung von Stichprobenkorrelationen vorsehen. Braeken und van Assen (2017) entwickelten eine eigenwertbasierte Methode, in der – ähnlich wie in NEST – Referenzniveaus für getestete Eigenwerte der Stichprobenkorrelationsmatrix definiert werden zur Bewertung des Kommunalitätsinkrementes durch einen Faktor. Den vorliegenden Ergebnissen zufolge ist es nicht trivial, ob und wann Pearson- bzw. polychorische Korrelationen

die Bestimmung der optimalen Faktorenzahl für kategoriale Variablen durch die Methode von Braeken und van Assen verbessern könnten.

Insgesamt wird die Verwendung von NEST zur Bestimmung der Faktorenzahl in explorativen Analysen durch die vorliegenden Ergebnisse unterstützt, wenngleich auf die verringerte Genauigkeit beider untersuchten NEST-Implementierungen gegenüber PA (siehe Simulation 3) für kategoriale Variablen mit heterogenen Item-Schwierigkeiten hingewiesen werden muss. Der Zuspruch für NEST in der vorliegenden Dissertation ist im Einklang mit den Arbeiten von Achim (2017, 2020), in denen NEST als präferierte Methode zur Bestimmung der Faktorenzahl bewertet wurde. In einer generelleren Betrachtung ist NEST als Test aufsteigender Faktorenzahlen anhand von Eigenwerten und deren simulierter Stichprobenverteilung eng verwandt mit der *Revised Parallel Analysis* zur Bestimmung der Faktorenzahl (RPA; Green et al., 2012)[19]. Der vorliegende Zuspruch für NEST ordnet sich daher auch im Einklang mit Zuspruch für RPA gegenüber PA in einen breiteren Literaturkontext ein (Braeken & van Assen, 2017; Green et al., 2012, 2015, 2016).

In einer gemeinsamen Diskussion von NEST und RPA steht der vorliegende Zuspruch für NEST im Widerspruch zu weiteren Simulationsstudien, in denen RPA eine niedrige Genauigkeit aufgrund häufiger Überschätzungen der optimalen Faktorenzahl zeigte (Auerswald & Moshagen, 2019; Cosemans et al., 2022; Lim & Jahng, 2019). Aufgrund der Ähnlichkeiten zwischen NEST und RPA sollten die Probleme von RPA auch in der vorliegenden Bewertung von NEST adressiert werden. Die entscheidende Gemeinsamkeit der kritischen Simulationsstudien zu RPA ist, dass RPA zu Überschätzungen neigte, wenn Populationen neben Faktoren zusätzlich durch zahlreiche sogenannte *Minor Factors* definiert wurden. Minor Factors bezeichnen diffuse Konstrukte, die nicht in einem reduktionistischen Faktorenmodell abgebildet werden können, aber den – komplexeren – datengenerierenden Prozess beeinflussen (Cattell, 1978; Fabrigar et al., 1999; Lorenzo-Seva et al., 2011; MacCallum et al., 2001; Preacher et al., 2013). In den vorliegenden Simulationen 1a bis 3 definierten ausschließlich die spezifizierten Faktorenmodelle die Populationen und keine zusätzlichen Minor Factors. Die

[19] NEST und RPA unterscheiden sich darin, dass NEST die Faktorenmodelle zur Simulation von Referenzdatensätzen anders als RPA berechnet und dass RPA die Bereinigung von Korrelationsmatrizen um ungeteilte Varianzen vor der Eigenwertzerlegung vorsieht (siehe Achim, 2017, für eine detaillierte Unterscheidung und einen simulationsbasierten Vergleich von NEST und RPA).

vorliegenden Ergebnisse zeigen daher nicht, in welchem Umfang NEST zu Überschätzungen der optimalen Faktorenzahl in Anwesenheit von Minor Factors neigt. Die Gemeinsamkeiten zwischen NEST und RPA legen jedoch nahe, dass NEST anfällig für Überschätzungen in Anwesenheit von Minor Factors ist (für eine Darstellung, wie sich Minor Factors auf getestete Eigenwerte in NEST auswirken und eine kritische Diskussion der qualitativen Unterscheidung zwischen Faktoren und Minor Factors in Simulationen, siehe Brandenburg & Papenberg, 2022, S. 22).

In einer kritischen Bewertung von NEST sollte im Einklang mit den kritischen Studien zu RPA (Auerswald & Moshagen, 2019; Cosemans et al., 2022; Lim & Jahng, 2019) festgehalten werden, dass die Signifikanz eines getesteten Eigenwertes in NEST nicht notwendigerweise einen theoretisch bedeutsamen Faktor anzeigt. Die vorliegenden Ergebnisse zeigen dieses Problem auch abseits von Minor Factors: In Simulation 3 produzierte $NEST_{Pearson}$ bei heterogenen Item-Schwierigkeiten häufiger signifikante Ergebnisse für Rauscheigenwerte als es das Signifikanzniveau (hier 5 %) vorgab. Ein ähnliches Problem ergibt sich aus den Unterschätzungen durch NEST in den durchgeführten Simulationen: Auch ein nicht-signifikanter Eigenwert in NEST impliziert nicht notwendigerweise die Abwesenheit eines bedeutsamen Faktors. Bei empirischen Anwendungen von NEST sollte daher insgesamt berücksichtigt werden, dass NEST die optimale Faktorenzahl sowohl unter- als auch überschätzen kann. Eine Präferenz für NEST gegenüber anderen Methoden kann vor dem Hintergrund der vorliegenden Ergebnisse dennoch begründet werden, da andere Methoden die gleichen Fehler oftmals mit höheren Wahrscheinlichkeiten zeigten.

Die beobachteten Verfehlungen der optimalen Faktorenzahl durch formale Methoden offenbaren, dass der Umgang mit dem Faktorenzahlproblem in explorativen Faktorenanalysen Richtlinien erfordert, die über die Wahl einer möglichst genauen Methode hinausgehen. Die Erkenntnisse über die Genauigkeit von Methoden aus Simulationsstudien tragen dazu bei, Verfehlungen der optimalen Faktorenzahl im Vorhinein zu vermeiden. Die vorliegenden Ergebnisse zeigen, dass Verfehlungen der optimalen Faktorenzahl weniger wahrscheinlich sind unter verschiedenen Bedingungen, die in empirischen Anwendungen sichtbar oder planbar sind. Die Erhebung großer Stichproben reduzierte die Wahrscheinlichkeit der Auslassung einflussreicher Faktoren durch NEST (siehe Simulation 3). Außerdem reduzierten polychorische Korrelationen und kategoriale Variablen in Referenzdatensätzen die Anfälligkeit für die Akzeptanz von Difficulty Factors durch NEST und erhöhten die Genauigkeit für binäre Variablen.

Eine generelle Richtlinie zum Umgang mit möglichen Verfehlungen der optimalen Faktorenzahl ist zudem, bei explorativen Faktorenanalysen unterschiedliche Faktorenzahlen zu explorieren. Eine Präferenz für ein Faktorenmodell kann sich dann daraus ergeben, dass die Modellparameter bei einer bestimmten Faktorenzahl besser mit domänenspezifischem Wissen vereinbar sind als die Parameter bei anderen Faktorenzahlen (Cattell, 1978; Lim & Jahng, 2019; Lorenzo-Seva et al., 2011). Die Verwendung von NEST ergänzt die Abwägung unterschiedlicher Faktorenzahlen um den Aspekt, dass die statistische Evidenz für das jeweilige Kommunalitätsinkrement der Faktoren zufallskritisch bewertet wird (Achim, 2017, 2020). Die häufige Akzeptanz von Difficulty Factors durch $NEST_{Pearson}$ in Simulation 3 deutet jedoch darauf hin, dass die statistische Evidenz für Faktoren in NEST durch die Simulation einer ungeeigneten Stichprobenverteilung getesteter Eigenwerte verzerrt sein kann (siehe Brandenburg, 2022, S. 26). Die Diskussion von NEST-Lösungen in empirischen Anwendungen erfordert daher weitere Forschung mit dem Ziel, die notwendigen Bedingungen zu präzisieren, unter denen eine angemessene Stichprobenverteilung getesteter Eigenwerte simuliert wird.

In einer abschließenden, zusammenfassenden Einordnung der Validierung der untersuchten Methoden wird die Anwendung von NEST durch die vorliegende Dissertation insgesamt unterstützt. Im Vergleich der Methoden ist die vorrangige Anwendung von NEST dadurch gerechtfertigt, dass NEST in den durchgeführten Simulationen genauer war als EGA und – in geringerem Umfang und nicht in allen Bedingungen (siehe oben) – als PA. Der Vergleich zwischen NEST und EGA wird durch die vorliegende Dissertation um den fundamentalen Unterschied ergänzt, dass NEST besser als EGA mit Kreuzladungen kompatibel ist, die in explorativen Faktorenanalysen nicht inhärent ausgeschlossen werden. Die Ergebnisse der durchgeführten Simulationen sind jedoch aufgrund der diskutierten Verfehlungen der optimalen Faktorenzahl kein Beleg dafür, dass NEST in allen empirischen Anwendungen optimale Lösungen produzieren würde. Die beobachteten Verfehlungen der optimalen Faktorenzahl durch alle untersuchten Methoden tragen zur Motivation bei, das Faktorenzahlproblem und Methoden zur Lösung weiter zu erforschen.

Literatur

Achim, A. (2017). Testing the number of required dimensions in exploratory factor analysis. *The Quantitative Methods for Psychology*, *13*(1), 64-74. https://doi.org/10.20982/tqmp.13.1.p064

Achim, A. (2020). Esprit et enjeux de l'analyse factorielle exploratoire. [Spirit and issues of exploratory factor analysis.] *The Quantitative Methods for Psychology*, *16*(4), 213-247. https://doi.org/10.20982/tqmp.16.4.p213

Acton, G. S., & Revelle, W. (2002). Interpersonal personality measures show circumplex structure based on new psychometric criteria. *Journal of Personality Assessment*, *79*(3), 446-471. https://doi.org/10.1207/S15327752JPA7903_04

Acton, G. S., & Revelle, W. (2004). Evaluation of ten psychometric criteria for circumplex structure. *Methods of Psychological Research - Online*, 9(1). 1-27.

Alden, L. E., Wiggins, J. S., & Pincus, A. L. (1990). Construction of circumplex scales for the Inventory of Interpersonal Problems. *Journal of Personality Assessment*, *55*(3-4), 521-536. https://doi.org/10.1080/00223891.1990.9674088

Auerswald, M., & Moshagen, M. (2019). How to determine the number of factors to retain in exploratory factor analysis: A comparison of extraction methods under realistic conditions. *Psychological Methods*, *24*(4), 468-491. https://doi.org/10.1037/met0000200

Blanken, T. F., Deserno, M. K., Dalege, J., Borsboom, D., Blanken, P., Kerkhof, G. A., & Cramer, A. O. (2018). The role of stabilizing and communicating symptoms given overlapping communities in psychopathology networks. *Scientific Reports*, *8*, 5854. https://doi.org/10.1038/s41598-018-24224-2

Blondel, V. D., Guillaume, J. L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, *2008*(10), P10008. https://doi.org/10.1088/1742-5468/2008/10/P10008

Boulesteix, A. L., Binder, H., Abrahamowicz, M., & Sauerbrei, W. (2017). On the necessity and design of studies comparing statistical methods. *Biometrical Journal.*, *60*(1), 216-218. https://doi.org/10.1002/bimj.201700129

Borsboom, D. (2017). A network theory of mental disorders. *World Psychiatry*, *16*(1), 5-13. https://doi.org/10.1002/wps.20375

Borsboom, D., & Cramer, A. O. (2013). Network analysis: an integrative approach to the structure of psychopathology. *Annual Review of Clinical Psychology*, *9*, 91-121. https://doi.org/10.1146/annurev-clinpsy-050212-185608

Borsboom, D., Cramer, A. O., Schmittmann, V. D., Epskamp, S., & Waldorp, L. J. (2011). The small world of psychopathology. *PloS one*, *6*(11), e27407. https://doi.org/10.1371/journal.pone.0027407

Bourdeaux, M. J., Ozer, D. J., Oltmanns, T. F., & Wright, A. G. C. (2018). Development and Validation of the circumplex scales of interpersonal problems. *Psychological Assessment*, *30*(5), 594–609. https://doi.org/10.1037/pas0000505

Braeken, J., & van Assen, M. A. (2017). An empirical Kaiser criterion. *Psychological Methods*, *22*(3), 450-466. https://doi.org/10.1037/met0000074

Brandenburg, N. (2022). Factor retention in ordered categorical variables: Benefits and costs of polychoric correlations in eigenvalue-based testing. *Manuscript submitted for publication*.

Brandenburg, N., & Papenberg, M. (2022). Reassessment of innovative methods to determine the number of factors: A simulation-based comparison of exploratory graph analysis and next eigenvalue sufficiency test. *Psychological Methods*. Advance online publication. https://doi.org/10.1037/met0000527

Browne, M. W. (1992). Circumplex models for correlation matrices. *Psychometrika*, *57*(4), 469-497. https://doi.org/10.1007/BF02294416

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, *1*(2), 245-276. https://doi.org/10.1207/s15327906mbr0102_10

Cattell, R. B. (1978). *The scientific use of factor analysis in the behavioral and life sciences*. Plenum Press

Cho, S. J., Li, F., & Bandalos, D. (2009). Accuracy of the parallel analysis procedure with polychoric correlations. *Educational and Psychological Measurement*, *69*(5), 748-759. https://doi.org/10.1177/0013164409332229

Christensen, A. P., Garrido, L. E., & Golino, H. (2020, August 25). Comparing community detection algorithms in psychological data: A Monte Carlo simulation. *arXiv preprint*. https://doi.org/10.31234/osf.io/hz89e

Clark, D. A., & Bowles, R. P. (2018). Model fit and item factor analysis: Overfactoring, underfactoring, and a program to guide interpretation. *Multivariate Behavioral Research*, *53*(4), 544-558. https://doi.org/10.1080/00273171.2018.1461058

Cosemans, T., Rosseel, Y., & Gelper, S. (2022). Exploratory Graph Analysis for Factor Retention: Simulation Results for Continuous and Binary Data. *Educational and Psychological Measurement*, *82*(5), 880-910. https://doi.org/10.1177/00131644211059089

Cramer, A. O., van der Sluis, S., Noordhof, A., Wichers, M., Geschwind, N., Aggen, S. H., ... & Borsboom, D. (2012). Dimensions of normal personality as networks in search of equilibrium: You can't like parties if you don't like people. *European Journal of Personality*, *26*(4), 414-431. https://doi.org/10.1002/per.1866

Cramer, A. O., Waldorp, L. J., van der Maas, H. L., & Borsboom, D. (2010). Comorbidity: A network perspective. *Behavioral and Brain Sciences*, *33*(2-3), 137-150. https://doi.org/10.1017/S0140525X09991567

Di Blas, L., Grassi, M., Luccio, R., & Momentè, S. (2012). Assessing the interpersonal circumplex model in late childhood: The Interpersonal Behavior Questionnaire for Children. *Assessment*, *19*(4), 421-441. https://doi.org/10.1177/1073191111401172

Eaton, N. R., Krueger, R. F., Markon, K. E., Keyes, K. M., Skodol, A. E., Wall, M., Hasin, D. S., & Grant, B. F. (2013). The structure and predictive validity of the internalizing disorders. *Journal of Abnormal Psychology, 122*(1), 86–92. https://doi.org/10.1037/a0029598

Epskamp, S., Maris, G. K., Waldorp, L. J., & Borsboom, D. (2018). Network psychometrics. *arXiv preprint*, page *arXiv:1609.02818v2*.

Epskamp, S., Rhemtulla, M., & Borsboom, D. (2017). Generalized network psychometrics: Combining network and latent variable models. *Psychometrika*, *82*(4), 904-927. https://doi.org/10.1007/s11336-017-9557-x

Fabrigar, L. R., Visser, P. S., & Browne, M. W. (1997). Conceptual and methodological issues in testing the circumplex structure of data in personality and social psychology. *Personality and Social Psychology Review*, *1*(3), 184-203. https://doi.org/10.1207/s15327957pspr0103_1

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, *4*(3), 272-299. https://doi.org/10.1037/1082-989X.4.3.272

Fava, J. L., & Velicer, W. F. (1992). The effects of overextraction on factor and component analysis. *Multivariate Behavioral Research*, *27*(3), 387-415. https://doi.org/10.1207/s15327906mbr2703_5

Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods*, *9*(4), 466–491. https://doi.org/10.1037/1082-989X.9.4.466

Forbush, K. T., & Watson, D. (2013). The structure of common and uncommon mental disorders. *Psychological Medicine*, *43*(1), 97-108. https://doi.org/10.1017/S0033291712001092

Fried, E. I., & Cramer, A. O. (2017). Moving forward: challenges and directions for psychopathological network theory and methodology. *Perspectives on Psychological Science*, *12*(6), 999-1020. https://doi.org/10.1177/1745691617705892

Garrido, L. E., Abad, F. J., & Ponsoda, V. (2013). A new look at Horn's parallel analysis with ordinal variables. *Psychological Methods*, *18*(4), 454-474. https://doi.org/10.1037/a0030005

Garrido, L. E., Abad, F. J., & Ponsoda, V. (2016). Are fit indices really fit to estimate the number of factors with categorical variables? Some cautionary findings via Monte Carlo simulation. *Psychological Methods*, *21*(1), 93-111. https://doi.org/10.1037/met0000064

Golino, H. F., & Epskamp, S. (2017). Exploratory graph analysis: A new approach for estimating the number of dimensions in psychological research. *PloS one*, *12*(6), e0174035. https://doi.org/10.1371/journal.pone.0174035

Golino, H., Shi, D., Christensen, A. P., Garrido, L. E., Nieto, M. D., Sadana, R., ... & Martinez-Molina, A. (2020). Investigating the performance of exploratory graph analysis and traditional techniques to identify the number of latent factors: A simulation and tutorial. *Psychological Methods, 25*(3), 292-320. https://doi.org/10.1037/met0000255

Goretzko, D., & Bühner, M. (2020). One model to rule them all? Using machine learning algorithms to determine the number of factors in exploratory factor analysis. *Psychological Methods*, *25*(6), 776–786. https://doi.org/10.1037/met0000262

Goretzko, D., & Bühner, M. (2022). Factor retention using machine learning with ordinal data. *Applied Psychological Measurement*, 46(5), 406-421. https://doi.org/10.1177/01466216221089345

Goretzko, D., Pham, T. T. H., & Bühner, M. (2021). Exploratory factor analysis: Current use, methodological developments and recommendations for good practice. *Current Psychology*, *40*(7), 3510-3521. https://doi.org/10.1007/s12144-019-00300-2

Green, S. B., Levy, R., Thompson, M. S., Lu, M., & Lo, W. J. (2012). A proposed solution to the problem with using completely random data to assess the number of factors with parallel analysis. *Educational and Psychological Measurement*, *72*(3), 357-374. https://doi.org/10.1177/0013164411422252

Green, S. B., Redell, N., Thompson, M. S., & Levy, R. (2016). Accuracy of revised and traditional parallel analyses for assessing dimensionality with binary data. *Educational and Psychological Measurement*, *76*(1), 5-21. https://doi.org/10.1177/0013164415581898

Green, S. B., Thompson, M. S., Levy, R., & Lo, W. J. (2015). Type I and type II error rates and overall accuracy of the revised parallel analysis method for determining the number of factors. *Educational and Psychological Measurement*, *75*(3), 428-457. https://doi.org/10.1177/0013164414546566

Gurtman, M. B., & Pincus, A. L. (2000). Interpersonal Adjective Scales: Confirmation of circumplex structure from multiple perspectives. *Personality and Social Psychology Bulletin*, *26*(3), 374-384. https://doi.org/10.1177/0146167200265009

Gurtman, M. B., & Pincus, A. L. (2003). The circumplex model: Methods and research applications. In J. A. Schinka & W. F. Velicer (Eds.), *Handbook of Psychology: Research Methods in Psychology*, Vol. 2, (pp. 407–428). John Wiley & Sons Inc.

Guttman, L. (1954a). Some necessary conditions for common-factor analysis. *Psychometrika*, *19*(2), 149-161. https://doi.org/10.1007/BF02289162

Guttman, L. (1954b). A new approach to factor analysis: The radex. In P. F. Lazarsfeld (Ed.), *Mathematical Thinking in the Social Sciences* (pp. 258-348). Columbia University Press.

Hallquist, M. N., Wright, A. G., & Molenaar, P. C. (2021). Problems with centrality measures in psychopathology symptom networks: Why network psychometrics cannot escape psychometric theory. *Multivariate Behavioral Research*, *56*(2), 199-223. https://doi.org/10.1080/00273171.2019.1640103

Henson, R. K., & Roberts, J. K. (2006). Use of exploratory factor analysis in published research: Common errors and some comment on improved practice. *Educational and Psychological Measurement*, *66*(3), 393-416. https://doi.org/10.1177/0013164405282485

Hopwood, C. J., Ansell, E. B., Pincus, A. L., Wright, A. G., Lukowitsky, M. R., & Roche, M. J. (2011). The circumplex structure of interpersonal sensitivities. *Journal of Personality*, *79*(4), 707-740. https://doi.org/10.1111/j.1467-6494.2011.00696.x

Hopwood, C. J., & Good, E. W. (2019). Structure and correlates of interpersonal problems and sensitivities. *Journal of Personality*, *87*(4), 843-855. https://doi.org/10.1111/jopy.12437

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*(2), 179-185. https://doi.org/10.1007/BF02289447

Isvoranu, A.-M., Epskamp, S., & Cheung, M. W. -L. (2021). Network models of posttraumatic stress disorder: A meta-analysis. *Journal of Abnormal Psychology, 130*(8), 841–861. https://doi.org/10.1037/abn0000704

Jin, S., & Yang-Wallentin, F. (2017). Asymptotic robustness study of the polychoric correlation estimation. *Psychometrika*, *82*(1), 67-85. https://doi.org/10.1007/s11336d016-9512-2

Jonason, P. K., & Webster, G. D. (2010). The dirty dozen: A concise measure of the dark triad. *Psychological Assessment, 22*(2), 420–432. https://doi.org/10.1037/a0019265

Jöreskog, K. G. (1966). Testing a simple structure hypothesis in factor analysis. *Psychometrika*, *31*(2), 165-178. https://doi.org/10.1007/BF02289505

Jöreskog, K. G. (1978). Structural analysis of covariance and correlation matrices. *Psychometrika*, *43*(4), 443-477. https://doi.org/10.1007/BF02293808

Keith, T. Z., Caemmerer, J. M., & Reynolds, M. R. (2016). Comparison of methods for factor extraction for cognitive test-like data: Which overfactor, which underfactor? *Intelligence*, *54*, 37-54. https://doi.org/10.1016/j.intell.2015.11.003

Kotov, R., Ruggero, C. J., Krueger, R. F., Watson, D., Yuan, Q., & Zimmerman, M. (2011). New dimensions in the quantitative classification of mental illness. *Archives of General Psychiatry*, *68*(10), 1003-1011. https://doi.org/10.1001/archgenpsychiatry.2011.107

Krueger, R. F., & Markon, K. E. (2006). Reinterpreting comorbidity: A model-based approach to understanding and classifying psychopathology. *Annual Review of Clinical Psychology*, *2*, 111-133. https://doi.org/10.1146/annurev.clinpsy.2.022305.095213

Li, Y., Wen, Z., Hau, K. T., Yuan, K. H., & Peng, Y. (2020). Effects of cross-loadings on determining the number of factors to retain. *Structural Equation Modeling: A Multidisciplinary Journal*, *27*(6), 841-863. https://doi.org/10.1080/10705511.2020.1745075

Lim, S., & Jahng, S. (2019). Determining the number of factors using parallel analysis and its recent variants. *Psychological Methods, 24*(4), 452–467. https://doi.org/10.1037/met0000230

Locke, K. D. (2000). Circumplex scales of interpersonal values: Reliability, validity, and applicability to interpersonal problems and personality disorders. *Journal of Personality Assessment*, *75*(2), 249-267. https://doi.org/10.1207/S15327752JPA7502_6

Lorenzo-Seva, U., Timmerman, M. E., & Kiers, H. A. (2011). The Hull method for selecting the number of common factors. *Multivariate Behavioral Research*, *46*(2), 340-364. https://doi.org/10.1080/00273171.2011.564527

Lubbe, D. (2019). Parallel analysis with categorical variables: Impact of category probability proportions on dimensionality assessment accuracy. *Psychological Methods*, *24*(3), 339–351. https://doi.org/10.1037/met0000171

MacCallum, R. C., Widaman, K. F., Preacher, K. J., & Hong, S. (2001). Sample size in factor analysis: The role of model error. *Multivariate Behavioral Research*, *36*(4), 611-637. https://doi.org/10.1207/S15327906MBR3604_06

Marsman, M., Borsboom, D., Kruis, J., Epskamp, S., van Bork, R., Waldorp, L. J., ... & Maris, G. (2018). An introduction to network psychometrics: Relating Ising network models to item response theory models. *Multivariate Behavioral Research*, *53*(1), 15-35. https://doi.org/10.1080/00273171.2017.1379379

McCrae, R. R., Zonderman, A. B., Costa, P. T., Jr., Bond, M. H., & Paunonen, S. V. (1996). Evaluating replicability of factors in the Revised NEO Personality Inventory: Confirmatory factor analysis versus Procrustes rotation. *Journal of Personality and Social Psychology*, *70*(3), 552–566. https://doi.org/10.1037/0022-3514.70.3.552

McDonald, R. P., & Ahlawat, K. S. (1974). Difficulty factors in binary data. *British Journal of Mathematical and Statistical Psychology*, *27*(1), 82-99. https://doi.org/10.1111/j.2044-8317.1974.tb00530.x

McNeish, D., & Wolf, M. G. (2020). Thinking twice about sum scores. *Behavior Research Methods*, *52*(6), 2287-2305. https://doi.org/10.3758/s13428-020-01398-0

Muthén, B. (1978). Contributions to factor analysis of dichotomous variables. *Psychometrika*, *43*(4), 551-560. https://doi.org/10.1007/BF02293813

Neal, Z. P., & Neal, J. W. (2021). Out of bounds? The boundary specification problem for centrality in psychological networks. *Psychological Methods.* Advance online publication. https://doi.org/10.1037/met0000426

Nießl, C., Herrmann, M., Wiedemann, C., Casalicchio, G., & Boulesteix, A. L. (2022). Over-optimism in benchmark studies and the multiplicity of design and analysis options when interpreting their results. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *12*(2), e1441. https://doi.org/10.1002/widm.1441

Olsson, U. (1979a). On the robustness of factor analysis against crude classification of the observations. *Multivariate Behavioral Research*, *14*(4), 485-500. https://doi.org/10.1207/s15327906mbr1404_7

Olsson, U. (1979b). Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika*, *44*(4), 443-460. https://doi.org/10.1007/BF02296207

Pons, P., & Latapy, M. (2006). Computing communities in large networks using random walks. *Journal of Graph Algorithms and Applications*, *10*(2), 191-218. https://doi.org/10.1007/11569596_31

Preacher, K. J., Zhang, G., Kim, C., & Mels, G. (2013). Choosing the optimal number of factors in exploratory factor analysis: A model selection perspective. *Multivariate Behavioral Research*, *48*(1), 28-56. https://doi.org/10.1080/00273171.2012.710386

Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research*, *47*(5), 667-696. https://doi.org/10.1080/00273171.2012.715555

Remington, N. A., Fabrigar, L. R., & Visser, P. S. (2000). Reexamining the circumplex model of affect. *Journal of Personality and Social Psychology*, *79*(2), 286-300. https://doi.org/10.1037//0022-3514.79.2.286

Revelle, W. (2022). *Psych: Procedures for Psychological, Psychometric, and Personality Research*. Retrieved from https://cran.r-project.org/package=psych

Revelle, W., & Rocklin, T. (1979). Very simple structure: An alternative procedure for estimating the optimal number of interpretable factors. *Multivariate Behavioral Research*, *14*(4), 403-414. https://doi.org/10.1207/s15327906mbr1404_2

Rogoza, R., Cieciuch, J., & Strus, W. (2021). A three-step procedure for analysis of circumplex models: An example of narcissism located within the circumplex of personality metatraits. *Personality and Individual Differences*, *169*, 109775. https://doi.org/10.1016/j.paid.2019.109775

Roznowski, M., Tucker, L. R., & Humphreys, L. G. (1991). Three approaches to determining the dimensionality of binary items. *Applied Psychological Measurement*, *15*(2), 109-127. https://doi.org/10.1177/014662169101500201

Ruscio, J., & Roche, B. (2012). Determining the number of factors to retain in an exploratory factor analysis using comparison data of known factorial structure. *Psychological Assessment*, *24*(2), 282-292. https://doi.org/10.1037/a0025697

Schmitt, T. A. (2011). Current methodological considerations in exploratory and confirmatory factor analysis. *Journal of Psychoeducational Assessment*, *29*(4), 304-321. https://doi.org/10.1177/0734282911406653

Sharp, C., Wright, A. G. C., Fowler, J. C., Frueh, B. C., Allen, J. G., Oldham, J., & Clark, L. A. (2015). The structure of personality pathology: Both general ('g') and specific ('s') factors? *Journal of Abnormal Psychology, 124*(2), 387–398. https://doi.org/10.1037/abn0000033

Steer, R. A., Ball, R., Ranieri, W. F., & Beck, A. T. (1999). Dimensions of the Beck Depression Inventory-II in clinically depressed outpatients. *Journal of Clinical Psychology*, *55*(1), 117-128. https://doi.org/10.1002/(SICI)1097-4679(199901)55:1<117::AID-JCLP12>3.0.CO;2-A

Thurstone, L. L. (1931). Multiple factor analysis. *Psychological Review, 38*(5), 406–427. https://doi.org/10.1037/h0069792

Timmerman, M. E., & Lorenzo-Seva, U. (2011). Dimensionality assessment of ordered polytomous items with parallel analysis. *Psychological Methods*, *16*(2), 209-220. https://doi.org/10.1037/a0023353

Tracey, T. J. (2000). Analysis of circumplex models. In H. E. A. Tinsley & S. D. Brown (Eds.), *Handbook of Applied Multivariate Statistics and Mathematical Modeling* (pp. 641-664). Academic Press. https://doi.org/10.1016/B978-012691360-6/50023-9

Tran, U. S., & Formann, A. K. (2009). Performance of parallel analysis in retrieving unidimensionality in the presence of binary data. *Educational and Psychological Measurement*, *69*(1), 50-61. https://doi.org/10.1177/0013164408318761

van der Maas, H. L., Dolan, C. V., Grasman, R. P., Wicherts, J. M., Huizenga, H. M., & Raijmakers, M. E. (2006). A dynamical model of general intelligence: the positive manifold of intelligence by mutualism. *Psychological Review*, *113*(4), 842-861. https://doi.org/10.1037/0033-295X.113.4.842

Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika*, *41*(3), 321-327. https://doi.org/10.1007/BF02293557

Velicer, W.F., Eaton, C.A., Fava, J.L. (2000). Construct Explication through Factor or Component Analysis: A Review and Evaluation of Alternative Procedures for Determining the Number of Factors or Components. In R.D. Goffin & E. Helmes (Eds.), *Problems and Solutions in Human Assessment* (pp 41-71). Springer. https://doi.org/10.1007/978-1-4615-4397-8_3

Watson, D. (2005). Rethinking the mood and anxiety disorders: A quantitative hierarchical model for DSM-V. *Journal of Abnormal Psychology*, *114*(4), 522–536. https://doi.org/10.1037/0021-843X.114.4.522

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of Personality and Social Psychology*, *54*(6), 1063-1070. https://doi.org/10.1037/0022-3514.54.6.1063

Weng, L. J., & Cheng, C. P. (2005). Parallel analysis with unidimensional binary data. *Educational and Psychological Measurement*, *65*(5), 697-716. https://doi.org/10.1177/0013164404273941

Weng, L. J., & Cheng, C. P. (2017). Is categorization of random data necessary for parallel analysis on Likert-type data? *Communications in Statistics-Simulation and Computation*, *46*(7), 5367-5377. https://doi.org/10.1080/03610918.2016.1154154

Widaman, K. F. (2018). On common factor and principal component representations of data: Implications for theory and for confirmatory replications. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*(6), 829-847. https://doi.org/10.1080/10705511.2018.1478730

Williams, D. R., & Rast, P. (2020). Back to the basics: Rethinking partial correlation network methodology. *British Journal of Mathematical and Statistical Psychology*, *73*(2), 187-212. https://doi.org/10.1111/bmsp.12173

Yang, Y., & Xia, Y. (2015). On the number of factors to retain in exploratory factor analysis for ordered categorical data. *Behavior Research Methods*, *47*(3), 756-772. https://doi.org/10.3758/s13428-014-0499-2

Ziegler, M., & Hagemann, D. (2015). Testing the unidimensionality of items. *European Journal of Psychological Assessment*, *31*(4), 231-237. https://doi.org/10.1027/1015-5759/a000309

Zimmermann, J., & Wright, A. G. (2017). Beyond description in interpersonal construct validation: Methodological advances in the circumplex structural summary approach. *Assessment*, *24*(1), 3-23. https://doi.org/10.1177/1073191115621795

Zwick, W. R., & Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin*, *99*(3), 432-442.

Anhang: Einzelarbeiten

Die vorliegende Dissertation umfasst zwei Manuskripte, die die dargelegten Simulationen 1a bis 3 enthalten. Für jedes Manuskript wird im Folgenden offengelegt, welche Personen an unterschiedlichen Schritten bei der Erstellung der Manuskripte beitrugen. Der überwiegende Teil der Arbeit lag bei beiden Manuskripten beim Erstautor des Manuskriptes.

Brandenburg, N., & Papenberg, M. (2022). Reassessment of innovative methods to determine the number of factors: A simulation-based comparison of exploratory graph analysis and next eigenvalue sufficiency test. *Psychological Methods.* Advance online publication. https://doi.org/10.1037/met0000527

Planung: Brandenburg, N., & Papenberg, M.
Programmierung: Brandenburg, N., & Papenberg, M.
Durchführung: Brandenburg, N., & Papenberg, M.
Auswertung: Brandenburg, N., & Papenberg, M.
Manuskript: Brandenburg, N., & Papenberg, M.

Der Quellcode der Simulationen inklusive aller verwendeten Methodenimplementierungen und die rohen Simulationsergebnisse sind verfügbar unter https://osf.io/f7rw2/.

Brandenburg, N. (2022). Factor retention in ordered categorical variables: Benefits and costs of polychoric correlations in eigenvalue-based testing. *Manuscript submitted for publication*.

Planung: Brandenburg, N.
Programmierung: Brandenburg, N.
Durchführung: Brandenburg, N.
Auswertung: Brandenburg, N.
Manuskript: Brandenburg, N.

Der Quellcode der Simulationen inklusive aller verwendeten Methodenimplementierungen und die rohen Simulationsergebnisse sind verfügbar unter https://osf.io/wb2ys/.

AMERICAN PSYCHOLOGICAL ASSOCIATION

Psychological Methods

ISSN: 1082-989X

https://doi.org/10.1037/met0000527

Reassessment of Innovative Methods to Determine the Number of Factors: A Simulation-Based Comparison of Exploratory Graph Analysis and Next Eigenvalue Sufficiency Test

Nils Brandenburg and Martin Papenberg
Department of Experimental Psychology, Heinrich Heine University Düsseldorf

Abstract

Next Eigenvalue Sufficiency Test (NEST; Achim, 2017) is a recently proposed method to determine the number of factors in exploratory factor analysis (EFA). NEST sequentially tests the null-hypothesis that k factors are sufficient to model correlations among observed variables. Another recent approach to detect factors is exploratory graph analysis (EGA; Golino & Epskamp, 2017), which rules the number of factors equal to the number of nonoverlapping communities in a graphical network model of observed correlations. We applied NEST and EGA to data sets under simulated factor models with known numbers of factors and scored their accuracy in retrieving this number. Specifically, we aimed to investigate the effects of cross-loadings on the performance of NEST and EGA. In the first study, we show that NEST and EGA performed less accurately in the presence of cross-loadings on two factors compared with factor models without cross-loadings: We observed that EGA was more sensitive to cross-loadings than NEST. In the second study, we compared NEST and EGA under simulated circumplex models in which variables showed cross-loadings on two factors. Study 2 magnified the differences between NEST and EGA in that NEST was generally able to detect factors in circumplex models while EGA preferred solutions that did not match the factors in circumplex models. In total, our studies indicate that the assumed correspondence between factors and nonoverlapping communities does not hold in the presence of substantial cross-loadings. We conclude that NEST is more in line with the concept of factors in factor models than EGA.

Translational Abstract

Exploratory factor analysis (EFA) is a method to develop hypotheses concerning common factors governing correlations among variables. This makes EFA a valuable instrument in various fields of psychology (such as test development). A key problem in EFA is to determine the optimal number of factors that fits observed correlations and keeps resulting models parsimonious. Contemporary research on this problem does not provide consensus on the optimal solution. Next Eigenvalue Sufficiency Test (NEST; Achim, 2017) and exploratory graph analysis (EGA; Golino & Epskamp, 2017) are recently proposed methods to approach this problem. Both were shown to determine accurately the number of factors in simulated factor models in which variables indicated one factor each. In our report, we compare NEST and EGA with simulated factor models in which each variable indicated multiple factors to varying degrees. These conditions suit validation of methods to detect factors because the premise of an unknown number of factors implies that one may not assume how many factors link to individual variables. We conducted two simulation studies: In Study 1, we show that methods detect factors less accurately when variables indicated multiple factors each and highlight that EGA suffered stronger than NEST. In Study 2, we simulated circumplex models—a particular class of factor models—and show that NEST achieved high accuracy while EGA was strikingly inaccurate. We discuss reasons for the methods' performances and argue that the signal that EGA detects is incongruent on a statistical level with the understanding of factors in factor analysis.

Keywords: exploratory factor analysis, dimensionality assessment, next eigenvalue sufficiency test, exploratory graph analysis, cross-loadings

Multivariate data sets are common in psychological research. As an example, the Positive and Negative Affect Schedule (PANAS; Watson et al., 1988) is a 20-item instrument of which 10 are assumed to measure positive affect and the other 10 to measure negative affect. More technically, this instrument is a combination of two subscales indicating two distinct constructs. The division of

Nils Brandenburg https://orcid.org/0000-0002-7497-9384
Martin Papenberg https://orcid.org/0000-0002-9900-4268
Data and analysis scripts can be retrieved from the Open Science Repository: https://osf.io/f7rw2/.
Correspondence concerning this article should be addressed to Nils Brandenburg, Department of Experimental Psychology, Heinrich Heine University Düsseldorf, Universitätsstraße 1, 40225 Düsseldorf, Germany. Email: nils.brandenburg@hhu.de

items into subscales is justified by the observation that item pairs within each subscale correlate substantially stronger than pairs across different subscales.

Factor analysis (FA) is a widely used statistical procedure that formalizes the analysis of common constructs in items. In FA, observed variables are projected onto a smaller set of latent variables which are referred to as common factors. Researchers may use FA to estimate which observed variables link to which common factors, assessing the nature of factors in turn. The statistical link between observed variables and factors is commonly referred to as factor loadings. Returning to the PANAS, items on the positive affect subscale show substantial loading on a common factor whereas items on the negative affect subscale load on a different factor. Given that clear-cut subscales can be a desirable property which can be assessed through FA, FA plays a major role in test development (Conway & Huffcutt, 2003; Henson & Roberts, 2006; Timmerman & Lorenzo-Seva, 2011; Ziegler & Hagemann, 2015).

Mathematically, FA can be described as a linear model of correlations among observed variables. Let p be the number of observed variables and m be the number of factors.[1] A fundamental equation in FA is:

$$\Sigma = \Lambda\Psi\Lambda^{T} + \Theta \quad (1)$$

Here, Σ is the $p \times p$ correlation matrix of observed variables, Λ is the $p \times m$ matrix of factor loadings, Ψ is the $m \times m$ matrix of interfactor correlations, and Θ is the $p \times p$ matrix that accounts for variance and correlation that is not explained by the parameters in $\Lambda\Psi\Lambda^{T}$ (i.e., the systematic variance that is unique to variables or measurement error). Factor loadings can be viewed as regression weights that determine how scores in observed variables are functions of factor scores. We refer readers unfamiliar with FA to Widaman (2018) for a comprehensive introduction to FA and its differences to principal component analysis (PCA).

Exploratory factor analysis (EFA) is a special case of FA that estimates factor loadings without constraining parameters to 0—as in "there is no link between the variable and the factor." This way, EFA is a suitable tool to assess a structure of subscales when none is hypothesized prior to the analysis (Auerswald & Moshagen, 2019; Ruscio & Roche, 2012; Ziegler & Hagemann, 2015).

The focus of the present report is a problem that arises in EFA when no specific factorial structure can be assumed underlying an analyzed data set: How many factors should be considered to begin with? We refer to this as the number-of-factors problem. Deciding on the number of factors is of decisive importance to the quality of results in EFA from a substantive point of view: Had the authors of the PANAS committed to a three-factor solution instead of a two-factor solution, understanding of the measured constructs and the selection of items could have turned out differently. Of course, not all possible numbers of factors are equally appropriate for an analyzed data set; some data sets may reflect more factors than others. In general, if a model contains too few factors, model-implied correlations may not fit observed correlations (Zwick & Velicer, 1986). On the other hand, if a model contains too many factors, model-implied correlations may fit observed correlations achieving fit at the cost of overcomplexity (Fabrigar et al., 1999; Fava & Velicer, 1992; Lorenzo-Seva et al., 2011). Therefore, the number-of-factors problem requires that the optimal number of factors is assessed soundly.

Despite a decades-long search for a procedure that reliably determines the optimal number of factors in a data set, this problem has consistently avoided consensus among researchers and remains a topic of debate in contemporary psychometric research (Achim, 2017; Auerswald & Moshagen, 2019; Braeken & Van Assen, 2017; Cosemans et al., 2021; Goretzko & Bühner, 2020; Golino et al., 2020; Keith et al., 2016; Preacher et al., 2013). This debate has motivated the development of new methods that were proposed as preferable approaches to the number-of-factors problem based on accurate performance in simulation studies (Achim, 2017; Braeken & Van Assen, 2017; Goretzko & Bühner, 2020; Golino & Epskamp, 2017; Green et al., 2012; Ruscio & Roche, 2012). Simulations provide means to judge the performance of methods to detect factors: Random data sets can be generated under factor models with specifically engineered parameters and known numbers of factors which then may or may not be recovered by investigated methods. Among the newly developed methods, Achim (2017) proposed the Next Eigenvalue Sufficiency Test (NEST) and reported results from a simulation in which NEST outperformed competing methods. Similarly, Golino and Epskamp (2017) pioneered a novel approach based on network psychometrics—refer to Borsboom (2017) for an introduction to network analysis in psychometrics—in exploratory graph analysis (EGA), which was also reported to have outperformed competing methods in a simulation (Golino et al., 2020).

Importantly, while NEST and EGA were both presented as accurate methods to determine the number of factors, their relative advantages over one another remain largely undiscussed. Building on this gap in literature, we conducted a simulation-based review of NEST and EGA. In two studies, we applied NEST and EGA as well as traditional methods—as benchmark—to data sets under simulated factor models. We designed both of our studies to simulate problems that NEST and EGA handle differently by design. This way, we discuss how and when the differences between NEST and EGA manifest in divergent results and if one of the methods can be favored over the other. In fact, our results contribute to the ongoing debate on factor-based methods and network-based methods (Epskamp, Rhemtulla, & Borsboom, 2017; Hallquist et al., 2021; Neal & Neal, 2021; Van Bork et al., 2017) as our simulations uncovered conditions under which network models of data sets clearly diverge from the factor models that we used to simulate data sets.

We focused on the performance of NEST and EGA in data sets under factor models in which variables showed substantial loadings on more than one factor. Multiple loadings per variable are commonly referred to as cross-loadings. These are especially appealing for a review of NEST and EGA as neither method was tested in the presence of large cross-loadings. Golino et al. (2020) addressed this limitation to their simulation on EGA and suggested to investigate EGA under conditions with large cross-loadings. In the following sections, we introduce NEST and EGA to point out

[1] Technically, FA separates common factors which inform multiple variables from unique factors which inform only one variable. The present work only concerns common factors. Therefore, we refer to common factors simply by the term "factors."

how cross-loadings challenge NEST and EGA. Then, we report two studies that each included several large-scale simulations.

In our first study, we varied the number of factors on which variables were free to load while controlling all other aspects of the models. While both methods performed less accurately in the presence of cross-loadings, EGA suffered notably stronger than NEST. In our second study, we focused on circumplex models of correlations in multivariate data sets (Acton & Revelle, 2004; Fabrigar et al., 1997; Gurtman, & Pincus, 2003; Guttman, 1954; Tracey, 2000; Zimmermann & Wright, 2017). While the number-of-factors problem is not prominent in literature on circumplex models, these models can be expressed as factor models with cross-loadings and—importantly—an unambiguous number of factors. Our simulations demonstrate a striking divergence of NEST and EGA in such problems with a clear advantage to NEST. We also applied NEST and EGA to empirical data sets for which fitting circumplex models had been reported and discuss how the solutions of NEST and EGA in these data sets reflect the results of our simulations.

Next Eigenvalue Sufficiency Test

NEST (Achim, 2017; Achim, 2020; Achim, 2021) is a nonparametric test to determine the number of factors that builds on previous work by Green et al. (2012) and Ruscio and Roche (2012). To understand the idea that eigenvalues can be used in what the author called a "sufficiency test," we briefly outline the meaning of eigenvalues in NEST and EFA. In general, the eigenvalues of concern are those that result from eigenvalue decomposition of sample correlation matrices. This decomposition is a key step in parameter estimation in EFA and PCA (Revelle, n.d.) We account for PCA in this section because NEST actually tests eigenvalues as they occur in PCA despite it being a method designed for application in EFA. Specifically, PCA involves eigenvalue decomposition of the full-rank sample correlation matrix while EFA involves eigenvalue decomposition of reduced correlation matrices that have estimates of each variable's nonunique variance on the main diagonal. This way, EFA separates common variance from unique variance—hence, the term for unique variances in Equation 1 (Widaman, 2018).

Eigenvalues can be sorted by descending magnitude. Then, the k^{th} eigenvalue indicates the increment in modeled variance summed across all observed variables that would be achieved by adding the k^{th} factor to a model of otherwise $k - 1$ factors. Therefore, the link between eigenvalues and EFA or PCA is that the sum of the k largest eigenvalues is equal to the sum of modeled variance in a k-factor model. In FA, the sum of modeled common variance in a variable is referred to as the variable's *communality*.

Individual eigenvalues indicate the increment in communalities across all variables that is modeled through the addition of the corresponding factors (Auerswald & Moshagen, 2019). NEST uses eigenvalues of sample correlation matrices to test if a k-factor model is sufficient to explain observed variance or if the model should account for the variance that would be modeled through $k + 1$ factors. Specifically, NEST uses eigenvalues of full-rank sample correlation matrices as proxies for eigenvalues that would be found in EFA; this can be justified as data sets generated under a k-factor model show k particularly large eigenvalues from the full-rank correlation matrix.

NEST applies a sequence of tests of the null-hypothesis that k factors—with k starting at 0—are sufficient, incrementing k by 1 if the null-hypothesis is rejected. The test statistic is the eigenvalue at index $k + 1$. At a given iteration of k, a k-factor model of the analyzed data set is estimated to sample a large number of surrogate data sets under the k-factor model.[2] Because the k-factor model is the true data-generating process of these surrogate data sets, eigenvalues at index $k + 1$ from their correlation matrices form a sampling distribution of the test statistic conditional on the null-hypothesis that k factors are indeed sufficient. The observed eigenvalue at index $k + 1$ from the analyzed data set is expected to rank high among this distribution if k factors are insufficient for a factor model of the analyzed data set. For a given α-level and a number of surrogate data sets j, the null-hypothesis is rejected if the rank of the tested eigenvalue among the simulated sampling distribution is greater than $(1 - \alpha)(j + 1)$—higher ranks indicating greater eigenvalues.

It can be predicted that the statistical power of the null-hypothesis test in NEST decreases when several variables show substantial cross-loadings. Following Equation 1, elements of the matrix $\Lambda\Psi\Lambda^{\mathrm{T}}$ can be written as:

$$\rho_{ij} = \sum_{1 \le k \le m} (\lambda_{ik}\lambda_{jk} + \sum_{\substack{1 \le k' \le m \\ k' \ne k}} (\lambda_{ik}\lambda_{jk'}\psi_{kk'})) \tag{2}$$

In turn, variables correlate if they load on a common factor or on correlated factors. If the data-generating process of an analyzed data set is a factor model with substantial cross-loadings, the pattern of eigenvalues is different to that of a data set under a factor model without cross-loadings. As an extreme example, if all variables have positive loadings on all factors, all variables correlate as in a factor model with one general factor. Consequently, when substantial cross-loadings emulate a general factor, the largest eigenvalue increases and the other eigenvalues which correspond to factors decrease.

When eigenvalues corresponding to factors decrease in the presence of cross-loadings, it can be predicted that NEST will perform less accurately if the eigenvalues become so small that they hardly emerge from those that do not correspond to factors. However, if all eigenvalues that correspond to factors remain sufficiently larger than eigenvalues that do not correspond to factors, NEST can be predicted to maintain accuracy even in the presence of cross-loadings. Cross-loadings per se do not challenge NEST as it does not assume explicitly the absence of cross-loadings; it is merely the shift in the weights of factors possibly resulting from cross-loadings that decreases the signal-to-noise ratio for tested eigenvalues. Exploration of conditions under which NEST is robust to cross-loadings and conditions under which it is not is an open problem that we tackled in our research.

[2] The difference between NEST and revised parallel analysis (Green et al., 2012) is that NEST estimates the k-factor model through iterative optimization of communalities (Achim, 2017) while revised parallel analysis plugs in squared multiple correlations without iterative optimization. In the simulation by Achim (2017), NEST outperformed revised parallel analysis.

Exploratory Graph Analysis

Instead of assessing factor models through eigenvalues, EGA determines the number of factors through means that do not touch explicitly upon the principles of FA. EGA stems from an approach that models correlations through a network of pairwise dependencies among the variables themselves (Borsboom & Cramer, 2013; Cramer et al., 2012; Marsman et al., 2018). This approach is referred to as network psychometrics (Epskamp et al., 2018). In network psychometrics, observed variables and dependencies among them are modeled as graphical networks. A graph consists of nodes and edges that interconnect the nodes, thus forming a network. Applying this architecture to psychometrics, nodes represent variables and weighted undirected edges represent pairwise dependencies (Epskamp, Waldorp, et al., 2017). Results of EGA can broadly be described as a function of edges in a graphical model of the analyzed data sets. Hence, EGA exclusively accounts for common variance among variables, similar to EFA.

Network models may appear similar or even equivalent to factor models from a statistical point of view (Golino & Epskamp, 2017). However, abandoning common factors as the source of correlation changes the view of the nature of the analyzed data. The key difference between factor models and network models is that factor models commonly assume independence among observed variables conditional on the factors. In contrast, dependencies among variables that cannot be explained by latent variables are the assumed source of correlation in network psychometrics (Cramer et al., 2010; Cramer et al., 2012; Fried & Cramer, 2017). In literature on network psychometrics, it is argued that there are data sets for which pairwise dependencies among variables may be more appealing than common factors from a substantive point of view. For instance, Fried and Cramer (2017) argue that insomnia causes fatigue and concentration deficits, hence, explaining correlation among these three variables better than depression as an overarching factor. In summary, network psychometrics model correlation without assuming factors at all (Hallquist et al., 2021).

There is, however, an assumed correspondence between network models and factor models; EGA exploits this correspondence to determine the number of factors: In factor models, a set of variables is intercorrelated when they load on a common factor; in network models, a set of variables is intercorrelated when they form a community of interconnected nodes. A prominent proposition in network psychometrics is that communities in a network correspond to factors (Cramer et al., 2012; Epskamp et al., 2012; Epskamp et al., 2018; van der Maas et al., 2006).

In accordance to this proposition, EGA determines the number of factors in a data set through two essential steps: First, a network model of pairwise dependencies among variables is established by estimating the inverted covariance matrix of the analyzed data set. Second, a community-detection algorithm that groups variables into nonoverlapping communities is applied to the network model to estimate the number of communities (Golino & Epskamp, 2017; Golino et al., 2020). Based on the assumed correspondence between communities and factors, EGA rules the optimal number of factors equal to the number of communities in the network model. Refer to Golino et al. (2020) for a technical introduction to the estimation of the network model and the detection of communities as well as a tutorial on EGA. Figure 1 provides an illustration of EGA with a data set under a simulated factor model: It shows a network model of the data set and indicates the communities detected by EGA. Large-scale simulations in literature showed that communities and factors indeed correspond well in data sets sampled under factor models in which variables load on one factor each (Golino et al., 2020; Hallquist et al., 2021).

Regarding cross-loadings, the grouping of variables into nonoverlapping communities implies a problem with EGA. When a variable is sampled under a factor model with substantial loading on more than one factor, assigning it to one community only may be a misrepresentation of its relations to factors. Simply put, assigning a variable that loads equally on two factors to one community can lead to misleading conclusions if one infers factorial structures from communities only. Of course, conclusions from network psychometrics do not rely exclusively on such partitions, but partitioning variables into communities is key in EGA to detect factors. In fact, EGA not only suggests a number of factors but also implies which variables indicate a common factor—similar to a loading pattern. NEST, on the other hand, only suggests a number of factors. Our work focused on solutions to the number-of-factors problem; hence, we do not discuss in detail the interpretation of EGA results for data sets beyond the number of factors, nor do we discuss possible implications of cross-loadings to other functions of graphical models (e.g., node-centrality measures).

Although large cross-loadings may cause problems for the partitions into nonoverlapping communities, this does not necessarily imply that solutions to the number-of-factors problem are suboptimal. Afterall, correlations among variables under such factor models concern not only the graphical network estimated in EGA but also the performance of community detection algorithms. Golino et al. (2020) simulated factor models in which variables loaded primarily on one factor each and sampled minor cross-loadings on all other factor from a normal distribution $N(.00, .05)$. While they reported strong performance of EGA in their simulation, it cannot be inferred how EGA would perform under conditions in which the assignment of variables to factors is not as clear-cut. For these reasons, it is necessary to investigate the sensitivity of EGA to cross-loadings in large-scale simulations.

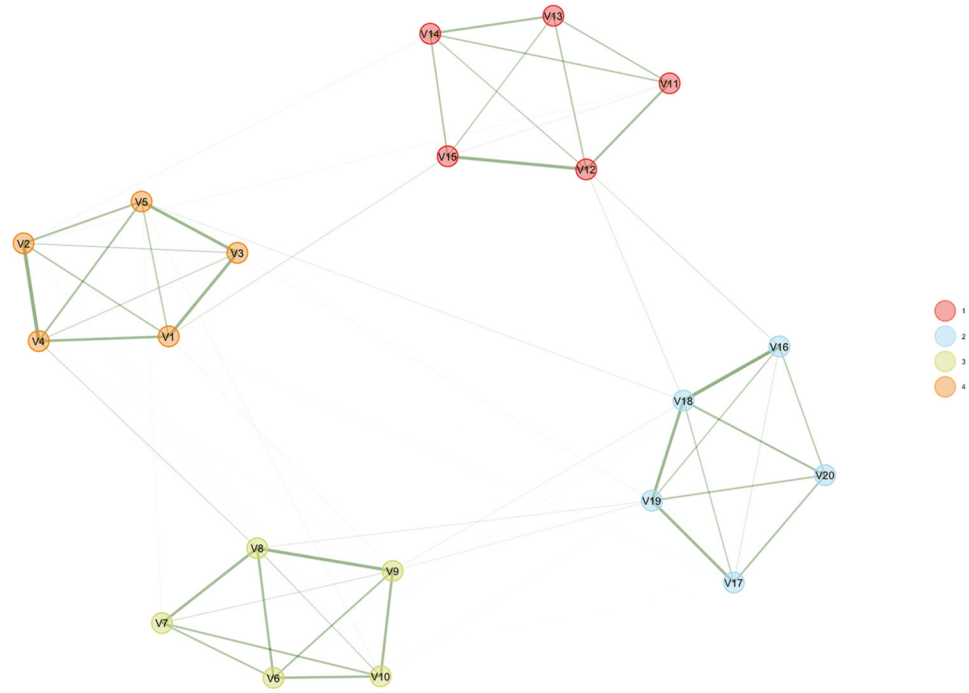
Study 1

Motivation

In our first study, we compared the performance of NEST and EGA in data sets under factor models without cross-loadings and under models with cross-loadings. We refer to factor models in which all variables load on one factor each as "independent-cluster models" and the number of nonzero loadings within a variable as "item complexity" (Revelle, n.d.).

As mentioned, NEST and EGA were both tested under conditions that match closely our definition for independent-cluster models (Achim, 2017; Golino et al., 2020). This is at odds with the problems that NEST, EGA, or EFA altogether are designed to solve. In general, lacking knowledge of the number of factors is incompatible with knowledge of item complexity. Of course, there are data sets for which one large loading per variable can be assumed (e.g., when variables have been selected to indicate one factor each). However, the method of choice for FA in such cases would be confirmatory factor analysis (CFA) if there is sufficient

Figure 1
Graphical Illustration of EGA



Note. This figure shows a graphical model of a data set of $N = 250$ sampled under a simulated factor model of four factors with five indicator variables each. All factor loadings were set to .70, all interfactor correlations were set to .30, and all cross-loadings were set to .00. Each node represents a variable from the data set. Nodes of the same color (shade) belong to the same community according to EGA. Line width indicates edge weights. The figure was created using the EGA function from the R package EGAnet (Version 1.0.0; Golino & Christensen, 2021). See the online article for the color version of this figure.

knowledge of the nature of the factors to justify this assumption in the first place (Achim, 2020). Literature provides examples of contexts in which cross-loadings are likely to be encountered: For instance, Acton and Revelle (2004) discussed how affective states can be modeled as linear combinations of both valence and arousal, and how some interpersonal traits are combinations of dominance and friendliness. Furthermore, Achim (2020) discussed that items in an exam on a subject may reflect both the "eagerness to study it" and the "ease of understanding" (p. 5) to varying degrees.

Method

Simulated Data

In Study 1, we conducted two simulations in which we sampled data sets under simulated factor models following a set of conditions. Both simulations shared a common design and differed only in simulated item complexity. The common design was a fully crossed design in which we manipulated five aspects of factor models and data sets: The number of factors (2, 3, 4, 5), the number of variables per factor (3, 4, 8, 12), the distribution of communalities in variables, $U(.40^2 - .05, .40^2 + .05)$; $U(.55^2 - .05, .55^2 + .05)$; $U(.70^2 - .05, .70^2 + .05)$; $U(.85^2 - .05, .85^2 + .05)$, the interfactor correlation (.00, .30, .50, .70), and the sample size of generated data sets (100, 250, 500, 1,000). For each of the 1,024 cells of this design, we generated 150 factor models and sampled one data set under each model, resulting in 153,600 data sets per simulation. In each data set, the number of variables was set to be the product of the number of factors and the number of variables per factor; the targeted communality of each variable was sampled from the according uniform distribution.

We adopted many of these manipulations from Golino et al. (2020). EGA showed strong performance in their simulation. We aimed for a design that included basic manipulations that are common in research on simulated factor models with the addition of manipulated item complexity. Our focus on item complexity required some alterations to the design from Golino et al. (2020). Notably, while Golino et al. (2020) manipulated loading parameters directly, we manipulated the variables' communalities. This was necessary because we aimed to vary item complexity between our simulations while keeping communalities on the same levels. We retained aspects of their design by sampling communalities from uniform distributions: The means of the described distributions are the squares of the means from the distributions of factor loadings from Golino et al. (2020)—hence, the expressions with squared values.

In the first of the two simulations in Study 1, we simulated independent-cluster models by forcing that all variables achieved their

targeted communality through one nonzero loading parameter. The number of variables that loaded on each factor was equal to the specified number of variables per factor from the common design.[3] Given that each variable only had one nonzero loading, the loading to achieve the targeted communality was the square root of the communality.

In the second simulation, we applied the same common design but induced item complexity through random allocation of communality across two factors in each variable. This way, simulated communalities remained the same as in the first simulation but the orientation of each variable relative to two assigned factors was free to vary anywhere between the factors. Each factor was indicated through a nonzero loading by a minimum number of variables equal to the number of variables per factor according to the simulation's design. We implemented this by assigning each variable to one factor as we did in the first simulation and then assigning each variable randomly to a second factor.

We devised a two-step procedure to determine which factor accounted for which amount of communality in each individual variable and to determine the according loading parameters. This procedure is not limited to simulation of two loadings per variable and generalizes to any degree of item complexity. First, for each variable, we sampled its targeted communality and determined the factors on which the variable loaded. Second, we distributed shares of the variable's communality across these factors. Sequentially, for each factor, we calculated the free communality h^2_{free} that had not yet been allocated to factors and sampled a random value from $U(0,\ h^2_{\text{free}})$ that determined the communality allocated to that factor. If the factor was the last one on which the variable loaded, the allocated communality was set so that the sum of communality shares across all factors amounted to the targeted communality of the variable. Updating h^2_{free} after each factor in the sequence implied diminishing expectations $E[U(0,\ h^2_{\text{free}})]$ at each step. Hence, we randomized the order in which factors were processed in this sequence within every variable to prevent that factors systematically accounted for more communality than other factors.

After distributing the targeted communality of a variable across the factors it loaded on, we determined the according loadings. This was done in a sequence across all factors that the variable loaded on. To explain how we solved this problem, Equation 2 can be rewritten to account for the communality of a variable (indexed by i):

$$h_i^2 = \sum_{1\le k\le m}\left(\lambda_{ik}^2 + \sum_{\substack{1\le k'\le m\\ k'\neq k}}(\lambda_{ik}\lambda_{ik'}\psi_{kk'})\right) \tag{3}$$

Here, m is the number of factors, k and k' are running indices for factors, and $\psi_{kk'}$ is the interfactor correlation between factors k and k'. For orthogonal factors (when all $\psi_{kk'}$ are 0), a variable's communality is the sum of its squared loadings. However, for correlated factors, the products of cross-loadings on correlated factors weighted by the respective interfactor correlation add to the communality. To isolate the contribution of one specific loading parameter to the communality, Equation 3 can be rewritten as:

$$h_i^2 = \sum_{1\le k<m}\left(\lambda_{ik}^2 + \sum_{\substack{1\le k'<m\\ k'\neq k}}(\lambda_{ik}\lambda_{ik'}\psi_{kk'})\right) + \lambda_{im}^2 + \sum_{1\le k<m}(2\lambda_{im}\lambda_{ik}\psi_{mk}) \tag{4}$$

Implementing the simulation, the problem at hand was to find the loadings to achieve the determined increment in communality factor-by-factor. Let Δ_{ik} denote the communality share of a factor indexed by k ($1 \le k \le m$) in variable i. Following Equation 4, Δ_{ik} can be expressed as:

$$\Delta_{ik} = \lambda_{ik}^2 + \sum_{1\le k'<k}(2\lambda_{ik}\lambda_{ik'}\psi_{kk'}) \tag{5}$$

For $k = 1$, the sum-term in Equation 5 does not apply and Equation 5 can be solved for λ_{ik} by $\lambda_{ik} = \sqrt{\Delta_{ik}}$. For $k > 1$, we solved Equation 5 for λ_{ik} by:

$$\lambda_{ik} = \frac{1}{2}\sum_{1\le k'<k}(2\lambda_{ik'}\psi_{kk'}) + \sqrt{\left(\frac{1}{2}\sum_{1\le k'<k}(2\lambda_{ik'}\psi_{kk'})\right)^2 + \Delta_{ik}} \tag{6}$$

For each factor model in both simulations, all interfactor correlations were set to the according value of the simulation design and all main-diagonal elements of the interfactor correlation matrix Ψ were set to 1. Finally, the matrix Θ was set to be a diagonal matrix that incremented all main-diagonal elements of $\Lambda\Psi\Lambda^{\mathrm{T}}$ to 1.

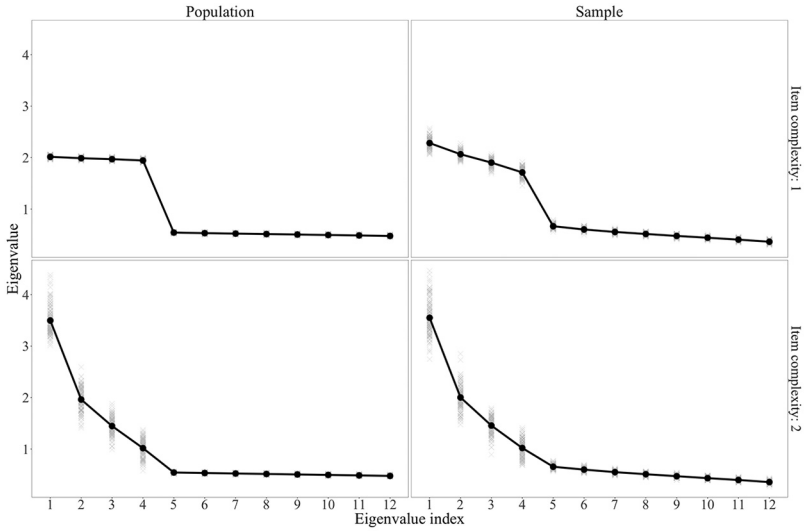
Figure 2 illustrates the effect of item complexity—as simulated in our work—on eigenvalues from correlation matrices. The increased magnitude of the largest eigenvalue is a result of the increased correlations among variables caused by cross-loadings. NEST can be predicted to perform worse under increasing complexity in all items because the last eigenvalue which corresponds to a factor decreases as a result of cross-loadings.

We used the statistical programming environment R (Version 4.1.0; R Core Team, 2021) to conduct all simulations in the present work. To simulate data sets under factor models, we determined all parameters in $\Lambda\Psi\Lambda^{\mathrm{T}} + \Theta$ and the implied correlation matrix Σ. Then, we sampled a data set with the according sample size under a multivariate normal distribution with all means set to 0 and the covariance matrix set to the implied Σ using the *rmvnorm* function from the R package *mvtnorm* (Version 1.1–3; Genz et al., 2021). The source code of our simulations and all results are available in the Open Science Framework repository that accompanies this article.

[3] For instance, when the number of factors was two and the number of variables per factor was three, three variables loaded on one factor and the rest loaded on the other factor.

Figure 2

Eigenvalues in Populations and Samples as a Function of Item Complexity



Note. This is an illustration how eigenvalues in simulated populations and samples from these populations changed depending on item complexity. The factor models in this figure were simulated with four factors, three variables per factor, targeted communalities in variables from $U(.70^2 - 0.05, .70^2 + 0.05)$, orthogonal factors, and a sample size of 250. In accordance to our simulations, we sampled 150 populations under these conditions and one sample per population. The gray crosses denote individual eigenvalues of populations and samples, the black dots connected by the black line indicate their mean at each index.

Method Implementations

In the present work, we used an implementation of NEST that was published by Achim (2017) as supplementary material. To test the eigenvalues of the analyzed sample correlation matrices at index $k + 1$, we generated 500 surrogate data sets under the respective k-factor model. All surrogate data sets matched the sample size and the number of variables of the analyzed data set and were sampled under multivariate normal distributions implied by the k-factor model with means set to 0.

We tested two variants of EGA that were available in the R package *EGAnet* (Version 1.0; Golino & Christensen, 2021). Both methods estimated a regularized partial-correlation network model—in network psychometrics commonly referred to as GLASSO estimation (Friedman et al., 2008)—but used different community-detection algorithms. Golino and Epskamp (2017) used the *Walktrap* algorithm (Pons & Latapy, 2006) in EGA but recent work (Christensen et al., 2020) showed that EGA also performed well with the *Louvain* algorithm (Blondel et al., 2008). We applied both variants using EGAnet with all settings at their default as of Version 1.0 and refer to them as $EGA_{Walktrap}$ and $EGA_{Louvain}$.

Given that our simulations included conditions under which neither NEST nor EGA had been tested before, we included variants of parallel analysis (PA; Horn, 1965) as a benchmark. In the previous studies on NEST (Achim, 2017) and EGA (Golino et al., 2020), PA also served as benchmark. Similar to NEST, PA compares eigenvalues of the analyzed sample correlation matrix with eigenvalues of surrogate data sets. In contrast to NEST, however, PA samples all surrogate data sets under a null-model that does not imply any underlying factors.

We applied PA implementations from the R package *psych* (Version 2.1.9; Revelle, 2022) and sampled 500 surrogate data sets per application through random permutation of all overserved variables from the analyzed data set. The package provides variants of PA using eigenvalue decomposition of full-rank correlation matrices or reduced correlation matrices. We included both variants in our simulations, referring to them as PA_{PCA} and PA_{FA} respectively. Additionally, psych provides estimation for reduced correlation matrices in PA_{FA} through a factor model with one factor (the default estimator in Version 2.1.9) or through squared multiple correlations of variables: We included both variants, referring to the former as $PA_{FA\text{-}1F}$ and the latter as $PA_{FA\text{-}SMC}$. Finally, we tested all variants of PA separately with the 95th percentile of reference eigenvalues as threshold for tested eigenvalues and with the 50th percentile. In total, we applied six variants of PA as benchmark for NEST and EGA: $PA_{PCA\text{-}95}$, $PA_{PCA\text{-}50}$, $PA_{FA\text{-}1F\text{-}95}$, $PA_{FA\text{-}1F\text{-}50}$, $PA_{FA\text{-}SMC\text{-}95}$, and $PA_{FA\text{-}SMC\text{-}50}$. For the sake of simplicity, we only report $PA_{PCA\text{-}50}$ and $PA_{FA\text{-}SMC\text{-}50}$ as representatives of PA_{PCA} and PA_{FA} in the following sections and note that no single variant of PA_{PCA} and PA_{FA} emerged as the most accurate nor the least accurate throughout our work (including Study 2). Appendix A provides an overview of results from Study 1 including all PA variants.

Data Analysis

We used the predetermined number of factors in each data set as ground truth to judge the performance of methods. Based on this ground truth, we categorized solutions (i.e., returned numbers of

factors) into four outcome categories: "accurate," "underestimation," "overestimation," and "undefined." We labeled solutions "accurate" if they were equal to the number of factors from the model under which the analyzed data set had been generated, as "underestimation" if it was smaller, and as "overestimation" if it was greater. In addition, we labeled solutions "undefined" if the method's implementation in R failed to converge on a solution (e.g., when a terminal error occurred during execution of the method).

Based on these labels, we defined the *accuracy* of a method as the proportion of accurate solutions in all recorded solutions. Note that data sets for which methods returned undefined solutions were included. Excluding undefined solutions and defining accuracy as the proportion of accurate solutions conditional on a method converging at all can result in misleading interpretations of accuracy. When the proportion of accurate solutions of a method is constant between two conditions but the proportion of undefined solutions changes, the conditional proportion of accurate solutions would change as well (Jaeger, 2008).

Results

Table 1 summarizes how methods performed in the simulations of Study. All methods performed considerably less accurate with increasing item complexity. NEST notably underestimated the number of factors in the presence of cross-loadings compared to independent-cluster models. The same applied to $PA_{PCA\text{-}50}$ and $PA_{FA\text{-}SMC\text{-}50}$, which is not surprising given that both are eigenvalue-based methods like NEST. $EGA_{Walktrap}$ and $EGA_{Louvain}$ suffered far stronger in the presence of cross-loadings than NEST and $PA_{FA\text{-}SMC\text{-}50}$. Compared to independent-cluster models, $EGA_{Walktrap}$ and $EGA_{Louvain}$ were more likely to underestimate and to overestimate the number of factors. While their accuracy was comparable to $PA_{PCA\text{-}50}$ in the presence of cross-loadings, NEST and $PA_{FA\text{-}SMC\text{-}50}$ were considerably more robust to cross-loadings.

NEST and both EGA variants failed to return a solution in many instances. To some extent, convergence failures occurred due to ill-behaved correlation matrices. In the second simulation, some conditions in which the number of factors was set to two produced model-implied correlation matrices Σ that were not positive semidefinite. This affected 26.74% of the population correlation matrices and 18.41% of the sample correlation matrices. In data sets with nonpositive-semidefinite sample correlation matrices, NEST failed to converge in 99.7% and both EGA variants failed to converge in 79.52%.

In Table 2, we isolated conditions with orthogonal factors as interfactor correlation and cross-loadings both—as implemented in our simulations—increased correlations among variables, essentially behaving like a simulated general factor. We specifically looked at orthogonal factors to isolate the effect of cross-loadings on the methods' performance. While cross-loadings still reduced accuracy in NEST and PA, the decline in EGA was again stronger.

Interim Discussion

The simulations in Study 1 indicate that NEST and EGA both performed less accurately in the presence of substantial cross-loadings on two factors than under independent-cluster models. For NEST, this decline could be predicted through examination of the effect of item complexity—as simulated in Study 1—on eigenvalues of correlation matrices. Compared to PA, NEST performed better than $PA_{PCA\text{-}50}$ and was generally on par with $PA_{FA\text{-}SMC\text{-}50}$ (see Table 1).

EGA only performed about as accurately as NEST and PA in the first simulation under independent-cluster models. Golino et al. (2020) suggested investigations into the performance of EGA in the presence of cross-loadings and our research indicates that cross-loadings indeed challenge EGA particularly strongly. Our results hint at an incompatibility between nonoverlapping communities with factors when factors are indicated by complex items. This incompatibility is highlighted by the increased frequencies of overestimations and underestimations by both EGA variants in the presence of cross-loadings: Hence, the decline in accuracy was not only caused by convergence failures but also due to partitions into communities that did not match the number of factors.

The results of the first simulation in Study 1 align well with the simulations by Achim (2017) and Golino et al. (2020) in that both NEST and EGA scored high accuracy in absence of substantial cross-loadings.

A limitation to Study 1 concerning ecological validity motivated us to conduct a second study with different factor models. In Study 1, we engineered factors models with the intention to investigate the effect of item complexity while controlling other aspects of the simulated models, including the communality of variables. Our design yielded large effects of item complexity, especially on EGA. As mentioned, factor models with substantial cross-loadings

Table 1
Outcomes per Method in Study 1

		Method				
Outcome	Item complexity	NEST	$EGA_{Walktrap}$	$EGA_{Louvain}$	$PA_{PCA\text{-}50}$	$PA_{FA\text{-}SMC\text{-}50}$
Accurate	1	84.2	79.2	79.8	75.6	83.3
	2	49.5	21.1	24.4	27.7	53.7
Underestimation	1	14.0	8.0	9.3	22.9	8.0
	2	41.2	53.6	50.0	72.0	41.7
Overestimation	1	1.5	1.2	1.4	1.5	8.7
	2	0.8	7.9	10.5	0.4	4.4
Undefined	1	0.2	11.6	9.5	0	0
	2	8.5	17.4	15.1	0	0.2

Note. The values denote percentage proportions of outcomes. $N = 307{,}200$.

Table 2
Accuracy per Method as a Function of the Number of Factors and Item Complexity

		Method				
No. of factors	Item complexity	NEST	$EGA_{Walktrap}$	$EGA_{Louvain}$	PA_{PCA-50}	$PA_{FA-SMC-50}$
2 factors	1	96.0	90.3	90.2	97.4	92.4
	2	42.7	23.2	15.2	47.4	74.1
3 factors	1	94.2	88.9	88.9	95.7	88.8
	2	80.2	46.8	53.7	68.9	82.6
4 factors	1	91.7	87.5	87.6	94.4	86.6
	2	78.3	30.0	29.6	67.3	80.0
5 factors	1	89.4	86.4	86.5	92.7	83.3
	2	75.3	25.5	23.0	62.7	77.1

Note. The values denote percentage proportions of accurate outcomes. Unlike Table 1, these results refer to conditions with uncorrelated factors only and do not distinguish different labels of inaccurate outcomes. N = 25,600.

are not uncommon in literature (Acton & Revelle, 2004; Hallquist et al., 2021; Kan et al., 2020; Rafaeli & Revelle, 2006; Reise, 2012; Revelle & Rocklin, 1979). However, the randomized loading patterns in our simulations are not necessarily representative of item complexity that occurs in empirical data sets. Consequently, it is questionable if the problems that occurred in EGA also occur in more ecologically valid simulations.

Study 2

Motivation

In Study 2, we intended to improve on ecological validity by targeting more common factor models that imply cross-loadings in many variables. A type of factor models that meets these criteria are circumplex models of correlations. The term "circumplex" in this context dates back to Guttman (1954); it describes correlation matrices in which correlations near the main-diagonal are high, then decrease with further distance from the main-diagonal, and then increase again near the boundaries of the matrix (Grassi et al., 2010). This can be illustrated as a circumplex pattern that maps variables as points on the circumference of a circle: The correlation between two variables decreases as the distance between the variables on the circumference increases. Circumplex models are common in different lines of psychological research. For instance, they are discussed as appropriate models of relations among affective states (Fabrigar et al., 1997; Gurtman & Pincus, 2000; Gurtman & Pincus, 2003; Remmington et al., 2000; Tracey, 2000) and interpersonal problems (Alden et al., 1990; Boudreaux et al., 2018; Di Blas et al., 2012; Hopwood et al., 2011; Hopwood & Good, 2019; Locke, 2000; Zimmermann & Wright, 2017).

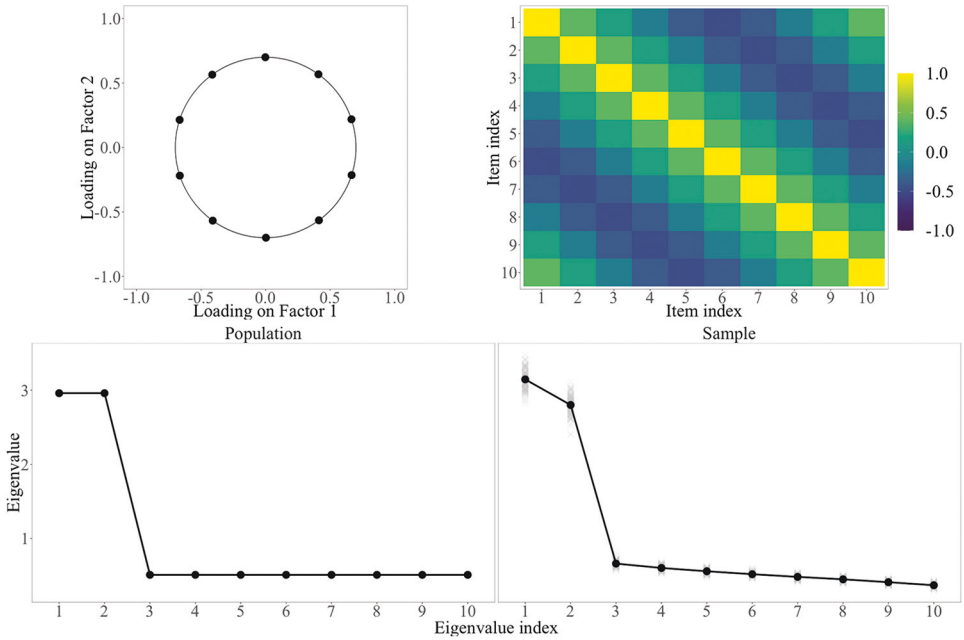
The link between factor models and circumplex models is that correlation matrices with the characteristic pattern of a circumplex model can be implied by factor models in which variables load on two orthogonal factors in a similar circular pattern: When the loadings of variables on two factors are treated as cartesian coordinates in a two-dimensional space, the described pattern of correlations emerges when these coordinates approximately form a circle around the origin (Acton & Revelle, 2002; Acton & Revelle, 2004; Fabrigar et al., 1997; Remmington et al., 2000). Obviously, this implies cross-loadings. Figure 3 illustrates how variables that load on two factors in a circumplex pattern show the characteristic correlations. Contrasting the eigenvalues from Study 1 (see Figure 2), circumplex models with two factors imply two large eigenvalues corresponding to the factors. It follows that NEST can be predicted to detect the two factors accurately. However, Figure 3 does not indicate predictions for the performance of EGA.

Circumplex models are not common in literature on the number-of-factors problem. One reason for this is that circumplex models are not necessarily entertained through FA (Browne, 1992). Another reason is that even when circumplex models are estimated through EFA or PCA, none of the formal methods to determine the number of factors are commonly applied (Alden et al., 1990; Boudreaux et al., 2018; Di Blas et al., 2012; Hopwood et al., 2011; Hopwood & Good, 2019; Locke, 2000). Still, circumplex models suffice our intention to review NEST and EGA in light of cross-loadings as they have a clearly defined number of factors. The simulated problem in Study 2 remains the same as in Study 1: If there is an optimal number of factors in generated data sets, can methods recover it?

We considered several types of circumplex models in Study 2 due to ambiguity in literature on circumplex models. First, the analyzed variables in circumplex models may be individual items (Alden et al., 1990; Boudreaux et al., 2018; Di Blas et al., 2012; Hopwood et al., 2011; Locke, 2000) or aggregations across subscales containing multiple items (Gurtman & Pincus, 2000; Hopwood & Good, 2019; Locke, 2000; Rogoza et al., 2021; Tracey, 2000; Zimmermann & Wright, 2017). Therefore, we designed different simulations to account for data sets under circumplex models on item-level and under circumplex-models on subscale-level.

Second, many definitions of circumplex models apply the constraint of equal spacing of modeled variables on the circumplex (Acton & Revelle, 2004; Gurtman & Pincus, 2000; Gurtman & Pincus, 2003; Wright et al., 2009). The simulated model in Figure 3 meets this constraint. However, the equal-spacing constraint may be lifted according to alternative definitions—although these models may then be referred to with different terms (Gurtman & Pincus, 2000; Rogoza et al., 2021). Grassi et al. (2010) argued that the equal-spacing constraint is overly restrictive in circumplex models on item-level. We agree with this position and designed our simulations accordingly: Our simulations of circumplex models on item-level did not force the models to meet the equal-spacing constraint;

Figure 3
Illustration of Circumplex Models



Note. The upper left panel shows factor loadings of 10 variables in a circumplex model. We set the communalities of all variables to .70. The upper right panel illustrates the circular pattern of correlations among variables: Correlations first decline with increasing distance to the main-diagonal and then increase again. The bottom half of the figure illustrates the eigenvalues of the model-implied correlation matrix on the left and the distribution of eigenvalues in 150 samples under this model on the right; the gray crosses denote individual eigenvalues and the black dots connected by the black line denote the sampling mean at each index. See the online article for the color version of this figure.

in our simulations of circumplex models on subscale-level, we varied whether the constraint applied or not.

Third, while definitions of circumplex models that relate to factor models generally focus on the circumplex pattern of loadings in a two-dimensional space, some definitions allow variables to load on an additional general factor (Acton & Revelle, 2004; Tracey, 2000). We account for the possibility that an additional general factor may be present in simulated circumplex models by conducting simulations with and without the general factor.

In summary, we designed six simulations in Study 2. In Simulations 1 and 2, we simulated circumplex models on item-level in which the individual variables themselves formed the circumplex pattern of loadings. We used two-dimensional factor models in Simulation 1 and three-dimensional factor models in Simulation 2, accounting for the general factor. In Simulations 3 and 4, we simulated circumplex models on subscale-level with eight distinct subscales. The models were implemented as multilevel factor models in which variables formed subscales through loadings on first-order factors—as in independent-cluster models—with interfactor correlations determined by a higher-order circumplex model. We simulated two-dimensional higher-order circumplex models on subscale-level in Simulation 3 and three-dimensional models in Simulation 4. In Simulations 5 and 6, we used the same data sets from Simulations 3 and 4 and computed sum scores across each subscale. This way, we accounted for applications that imply circumplex models on subscale-level in which aggregations across subscales (i.e., sum scores) are the analyzed variables.

General Method

The six simulations in Study 2 did not share a common design; hence, we report them separately. First, however, we explain methodological considerations that applied to all simulations. As in Study 1, we conducted all simulations using R (Version 4.1.0) and the mvtnorm package (Version 1.1–3) to sample data sets under multivariate normal distributions in accordance to simulated models. All six simulations followed a specific fully crossed design with a given number of conditions. We simulated 150 factor models per condition and one data set per factor model as we did in Study 1.

We applied the same implementations of NEST, EGA, and PA as in Study 1 and report $PA_{PCA\text{-}50}$ and $PA_{FA\text{-}SMC\text{-}50}$ as benchmark. Appendix B provides results of all PA variants in all simulation of Study 2. Furthermore, we retained the categorization of solutions into four outcome labels: accurate, underestimation, overestimation, and undefined.

Simulation 1

Simulated Data

In Simulation 1, we simulated circumplex models implied by factor models with two factors. As mentioned, we lifted the equal-spacing constraint in the simulated circumplex models. However, we divided the circumference of the circumplex into eight equally

sized parts (i.e., octants) and placed a specified number of variables into random spots within each part to ensure that variables were located all around the circumplex. We specifically decided to account for octants in our simulations in Study 2 because it is common to mark octants in circumplex models; octants may be used to assign substantive meaning to ranges around the circumference of the circumplex (Acton & Revelle, 2002; Boudreaux et al., 2018; Gurtman & Pincus, 2003; Hopwood & Good, 2019) and to group variables within octants to subscales (Alden et al., 1990; Locke, 2000; Zimmermann & Wright, 2017). Given that all simulated circumplex models in Simulation 1 included two factors, we set the optimal number of factors to two in all data sets.

We manipulated three aspects of the circumplex models and data sets in a fully crossed design: The number of variables per octant (3, 6, 8, 12), the targeted communalities of variables, $U(.40^2 - .05, .40^2 + .05)$; $U(.55^2 - .05, .55^2 + .05)$; $U(.70^2 - .05, .70^2 + .05)$; $U(.85^2 - .05, .85^2 + .05)$,[4] and the sample size (100, 250, 500, 1,000).

We devised a two-step procedure to determine the loadings of variables with respect to their targeted communality and their respective octant in the circumplex. First, for each variable in a model, we sampled its communality h^2 under the according uniform distribution. Then, we defined the vector $(\sqrt{h^2}, 0)$, which served as a reference vector of loadings that would achieve the variable's targeted communality.[5] Relative to the vector (1, 0), the centers of the octants were located at angles of 0°, 45°, 90°, 135°, 180°, 225°, 270°, and 315°; all octants spanned a range of 45° on the circumplex. We rotated the vector $(\sqrt{h^2}, 0)$ counterclockwise by the sum of the angle of the according octant and a random value under $U(-\frac{45°}{2}, \frac{45°}{2})$. Figure 4 illustrates the loading pattern, the implied correlations, and eigenvalues of a circumplex model in accordance to Simulation 1.

Results

Table 3 summarizes the performance of all applied methods. As predicted, NEST detected two factors accurately because the two largest eigenvalues were sufficiently larger than all other eigenvalues. This pattern also resulted in near-perfect accuracy in $PA_{PCA\text{-}50}$ and $PA_{FA\text{-}SMC\text{-}50}$. $EGA_{Walktrap}$ and $EGA_{Louvain}$ detected more than two factors in most of the data sets. In addition, both variants frequently failed to return solutions. Figure 5 provides a more detailed account of the solutions by each method as a function of the number of variables per octant. The figure indicates that both EGA variants deviated from two-factor solutions with an increasing number of variables per octant and instead favored three- and four-factor solutions.

The contrast between the eigenvalue-based methods and EGA in Simulation 1 from Study 2 was far greater than in Study 1. This adds support to the position that EGA is often unable to determine accurately the number of factors through nonoverlapping communities if variables show substantial cross-loadings.

Simulation 2

Simulated Data

In Simulation 2, we used the same numbers of variables per octant, targeted communalities, and sample sizes as in Simulation 1. We extended the design through the addition of a general factor to the circumplex models. To this end, we varied the share of the variables' targeted communalities that was achieved through loadings on the two factors that constituted the circumplex (3/6, 4/6, 5/6); these three levels were added to the fully crossed design. When the two circumplex factors in the model accounted for 4/6 of the communality in each variable, all three factors each accounted the same share of communality across all variables; the general factor was more pronounced than the circumplex factors when their communality share was 3/6 and it was less pronounced when their share was 5/6.

We used the same two-step procedure to derive factor loadings as in Simulation 1: For each variable, we first sampled its communality h^2. For a given communality share q of the two circumplex factors, we then rotated the reference vector $(\sqrt{qh^2}, 0)$ as in Simulation 1. The loading on the general factor was set to $\sqrt{(1-q)h^2}$.

Similar to Figure 4 from Simulation 1, Figure 6 illustrates the simulated circumplex models in accordance to the conditions of Simulation 2. The presence of the general factor increased the optimal number of factors to three.

Results

Table 4 indicates the same pattern of results as in Simulation 1. Again, NEST and the PA variants were accurate while $EGA_{Walktrap}$ and $EGA_{Louvain}$ frequently detected more than three factors and suffered from convergence failures. We also provide a detailed account of solutions as a function of the number of variables per octant in Figure 7. Similar to Simulation 1, EGA was prone to overestimation with an increasing number of variables per octant. Both EGA variants preferred solutions with four or five factors.

The results are similar to Simulation 1 and further indicate incompatibility between EGA and circumplex models. EGA appeared sensitive to the presence of a general factor in the circumplex models given its preference for four- or five-factor solutions (see Figure 7) whereas EGA preferred three- or four-factor solutions without the general factor (see Figure 5).

Simulation 3

Simulated Data

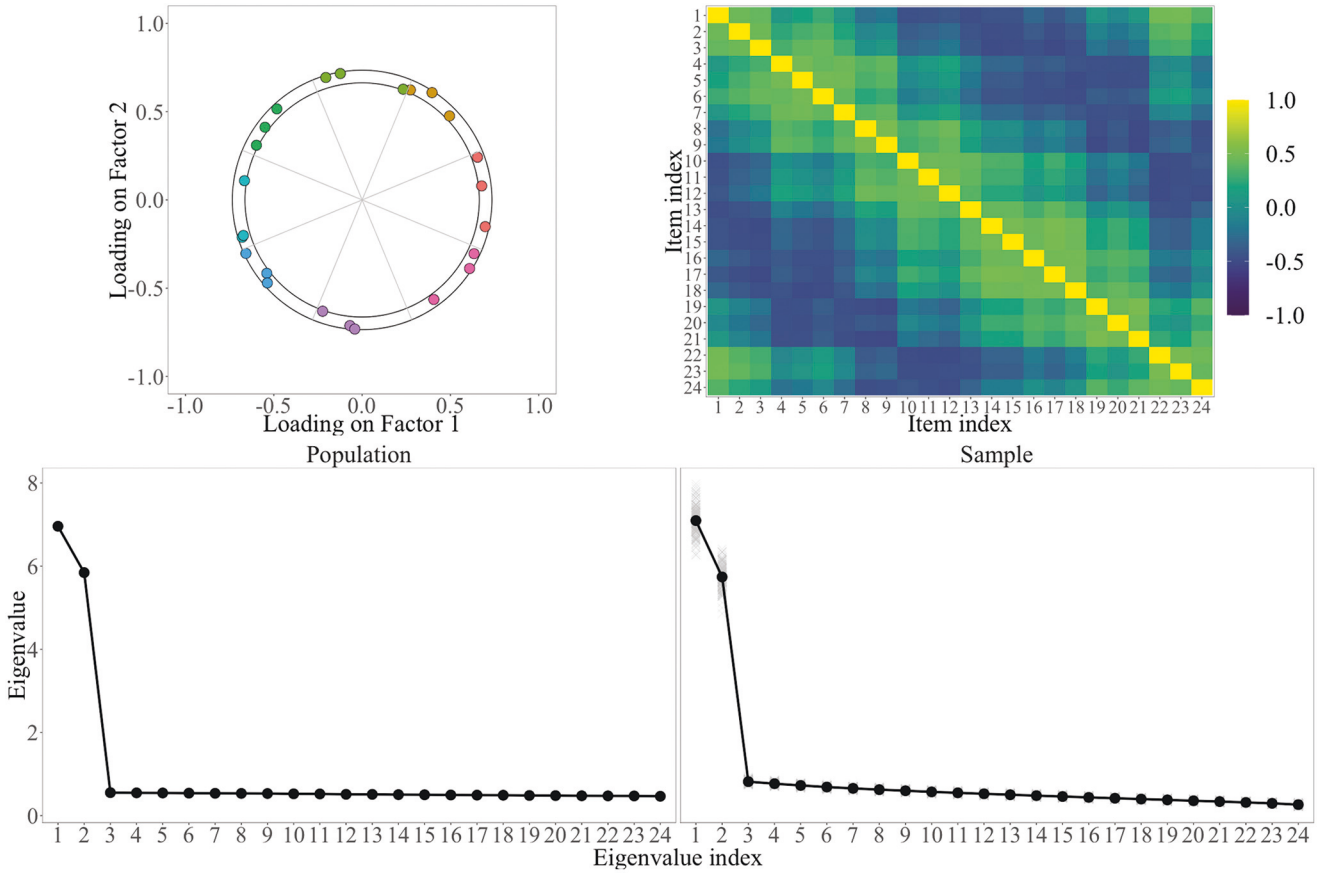
In Simulation 3, we simulated data sets under models that resembled items distributed across distinct subscales that indicate factors which, in turn, correlate in accordance to a circumplex model on subscale-level. This was implemented through multilevel factor models in which variables loaded on one first-order factor each and the first-order factors loaded on higher-order factors in a circumplex model on subscale-level. The fully crossed simulation design was more complex than in Simulations 1 and 2.

We manipulated the number of variables per subscale (3, 6, 8, 12) so that all subscales contained the same number of variables. Moreover, we manipulated the communalities of variables in the independent-cluster models on item-level, $U(.40^2 - .05, .40^2 + .05)$; $U(.55^2 - .05, .55^2 + .05)$; $U(.70^2 - .05, .70^2 + .05)$; $U(.85^2 - .05, .85^2$

[4] These distributions are the same as in Study 1.

[5] We did not need to account for interfactor correlation as factors were always orthogonal in simulated circumplex models.

Figure 4
Illustration Circumplex Models in Simulation 1



Note. The upper left panel shows factor loadings in a randomly generated circumplex model. We simulated three variables per octant; dots of a common color (shade) denote variables of a common octant, which are marked by the gray lines. We sampled all communalities from $U(.70^2 - 0.05, .70^2 + 0.05)$. The range of possible communalities is indicated by the distance between the two solid black circles. The upper right panel illustrates the model-implied correlations among variables. The bottom half of the figure illustrates the eigenvalues of the model-implied population correlation matrix on the left and the distribution of eigenvalues in 150 samples under this model with $N = 250$ on the right; the gray crosses denote individual eigenvalues and the black dots connected by the black line denote the sampling mean at each index. See the online article for the color version of this figure.

+ .05), and the communalities of first-order factors in the circumplex models on subscale-level, $U(.40^2 - .05, .40^2 + .05)$; $U(.55^2 - .05, .55^2 + .05)$; $U(.70^2 - .05, .70^2 + .05)$; $U(.85^2 - .05, .85^2 + .05)$. The number of first-order factors was set to eight in all simulated models. Given that all variables loaded on one first-order factor only, the loading to achieve the targeted communality was always the square root of that communality. The circumplex on subscale-level was divided into octants as in Simulation 1 using the same central orientations. Each subscale corresponded to a unique octant. In contrast to Simulation 1, we manipulated if the circumplex models met the equal-spacing constraint by specifying subscale-specific angular deviations from the central orientation of its respective octant: setting it to 0; sampling it from $U(-\frac{45°}{2}, \frac{45°}{2})$. Finally, we manipulated the sample size of data sets (100, 250, 500, 1,000).

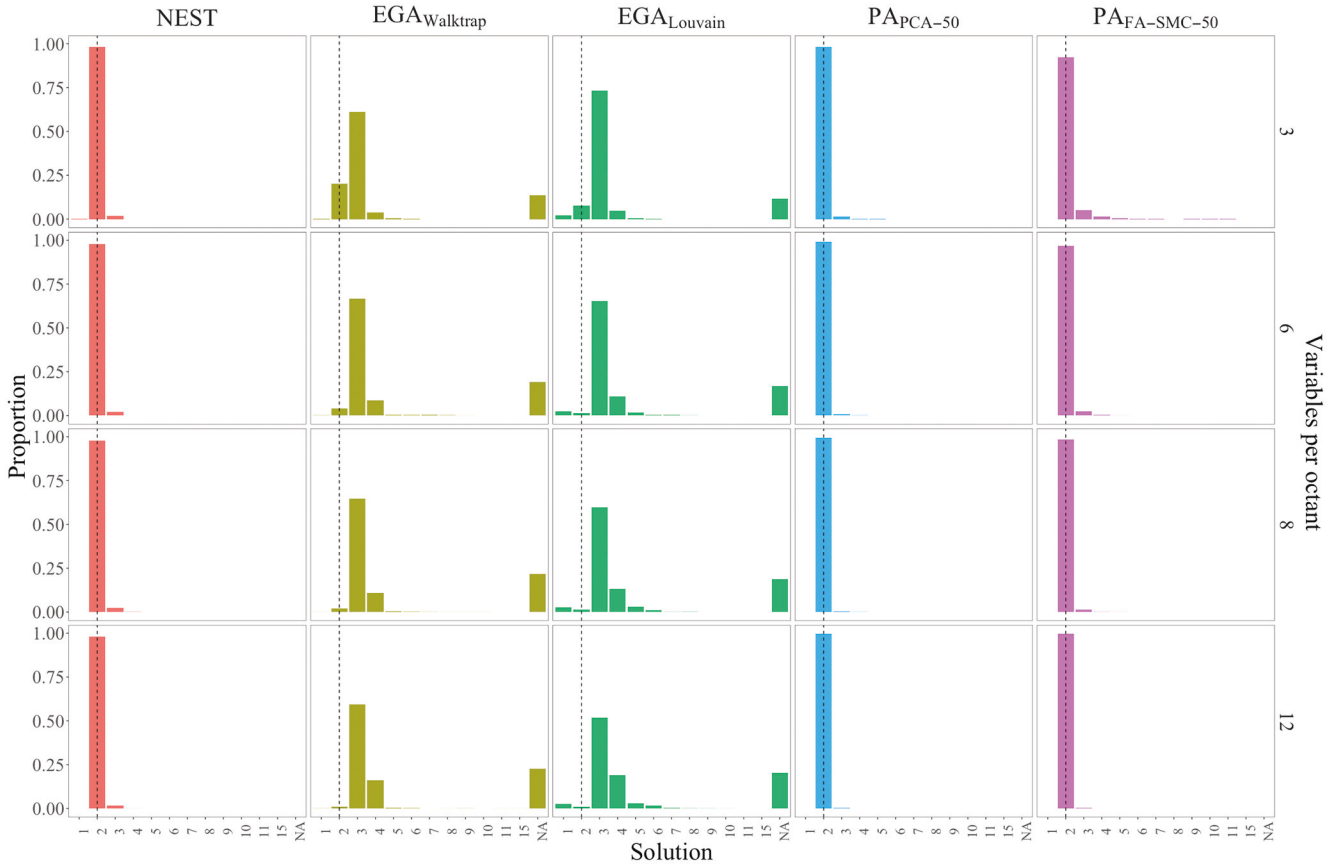
Table 3
Outcomes per Method in Simulation 1 of Study 2

	Method				
Outcome	NEST	$EGA_{Walktrap}$	$EGA_{Louvain}$	PA_{PCA-50}	$PA_{FA-SMC-50}$
Accurate	98.0	6.9	2.9	99.2	96.8
Underestimation	0	0.1	2.6	0	0
Overestimation	2.0	73.6	77.7	0.8	3.2
Undefined	0	19.3	16.9	0	0

Note. The values denote percentage proportions of outcomes. $N = 9{,}600$.

Figure 5

Solutions per Method as a Function of Variables per Octant in Simulation 1



Note. The vertical dashed line in each column marks the number of factors in the simulated circumplex models (two). The x-axis is discontinuous because it lists only solutions that were returned at least once by one method during the simulation. Undefined estimates are scored above the rightmost label (NA). N = 9,600. See the online article for the color version of this figure.

Figure 8 illustrates an example of the simulated models from Simulation 3. Eight eigenvalues emerge in Figure 8—corresponding to the eight first-order factors—but the largest two eigenvalues are considerably larger than the following six, reflecting the circumplex model on subscale-level. Similar to the models in Study 1, the circumplex model on subscale-level implied interfactor correlations, which, in turn, shifted weight toward the factors that modeled the interfactor correlations. Nevertheless, we set the optimal number of factors to eight.

Results

Table 5 summarizes the performance of methods in Simulation 3. Here, all methods showed similar accuracy. NEST and the PA variants were prone to detect less than eight factors; both EGA variants also tended to detect less than eight factors but again suffered from convergence failures.

Figure 9 shows solutions as a function of the targeted communality of first-order factors in the circumplex model on subscale-level. Higher communality in the circumplex model implied a greater shift in weight toward the higher-order factors corresponding to the circumplex model. When the factors of the circumplex model were most pronounced, NEST and PA$_{PCA-50}$ were biased toward two-factor solutions. PA$_{FA-SMC-50}$ was more liberal than NEST and PA$_{PCA-50}$ and, therefore, also favored two-factor solutions but to a lesser extent. Importantly, EGA$_{Walktrap}$ and EGA$_{Louvain}$ did not detect the number of factors according to the circumplex model when it was most pronounced; instead, both variants indicated a tendency toward four-factor solutions.

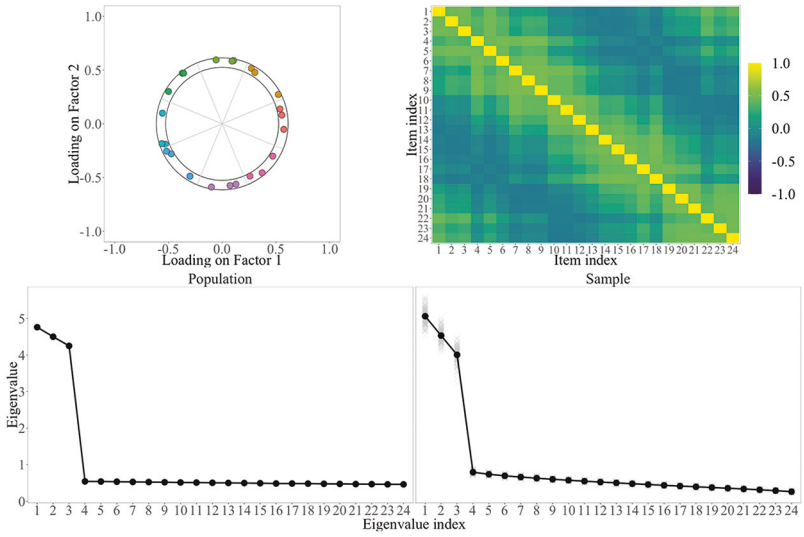
In summary, the gradual manipulation of communalities in the circumplex models on subscale-level showed that all methods were comparably accurate when the independent-cluster model on item-level was less overshadowed by the higher-order circumplex model. The preference for the number of factors from the circumplex models observed in NEST and PA when it was indicated more strongly may be considered reasonable; similar to Simulation 1 and 2, EGA did not detect the factors from the circumplex models under such conditions.

Simulation 4

Simulated Data

In Simulation 4, we applied the same design as in Simulation 3 but added a general factor to the higher-order circumplex models on subscale-level. Similar to Simulation 2, we allocated the communalities of subscales in the circumplex models across two circumplex

Figure 6
Illustration Circumplex Models as Simulated in Simulation 2



Note. This figure is the pendant to Figure 5 for Simulation 2. The circumplex model that is illustrated in this figure was simulated under the same conditions as Figure 5 with the exception of a general factor indicated by all variables. Hence, the smaller radius in the upper left panel, the higher correlations among variables compared to Figure 5, and the additional emergent eigenvalue compared to Figure 5. The loadings on the circumplex factors accounted for 4/6 of the communality in each variable. See the online article for the color version of this figure.

factors and the general factor. Given that the design of Simulation 4 implied greater computational effort than Simulation 2, we decided to weigh all three higher-order factors in the circumplex models equal and set the communality share of the two circumplex factors to 4/6.

Figure 10 illustrates a model in accordance to the conditions of Simulation 4. While the number of factors in the circumplex models was different than in Simulation 3, all independent-cluster models on item-level included eight factors to simulate eight subscales. Hence, we set the optimal number of factors to eight as in Simulation 3.

Results

Table 6 summarizes the results from Simulation 4 and is consistent with Simulation 3. All methods detected less than eight factors in many instances and both EGA variants again suffered from convergence failures. Figure 11 shows solutions as a function of communalities in the circumplex models on subscale-level. Again, NEST and $PA_{PCA\text{-}50}$ favored the number of factors from the circumplex model when it was most pronounced. $PA_{FA\text{-}SMC\text{-}50}$ was again more liberal and hence less biased toward three-factor solutions. While both EGA variants were biased toward four-factor solutions in Simulation 3—without the general factor—both were biased toward five-factor solutions in the presence of the general factor in the circumplex models.

Results from Simulation 4 are in line with Simulation 3 as all methods scored similar accuracy when the independent-cluster model on item-level was not overshadowed by the higher-order circumplex model. Otherwise, NEST and PA tended toward the number of factors from the circumplex model while EGA did not. Similar to Simulations 1 and 2, EGA indicated sensitivity to the presence of a general factor in circumplex models as it often detected an additional factor in Simulation 4 compared to Simulation 3.

Simulation 5

Simulated Data

In Simulation 5, we did not generate new data sets and instead used the 76,800 data sets from Simulation 3. The difference to Simulation 3 was that we computed sum scores across the subscales indicated by the independent-cluster models on item-level and then

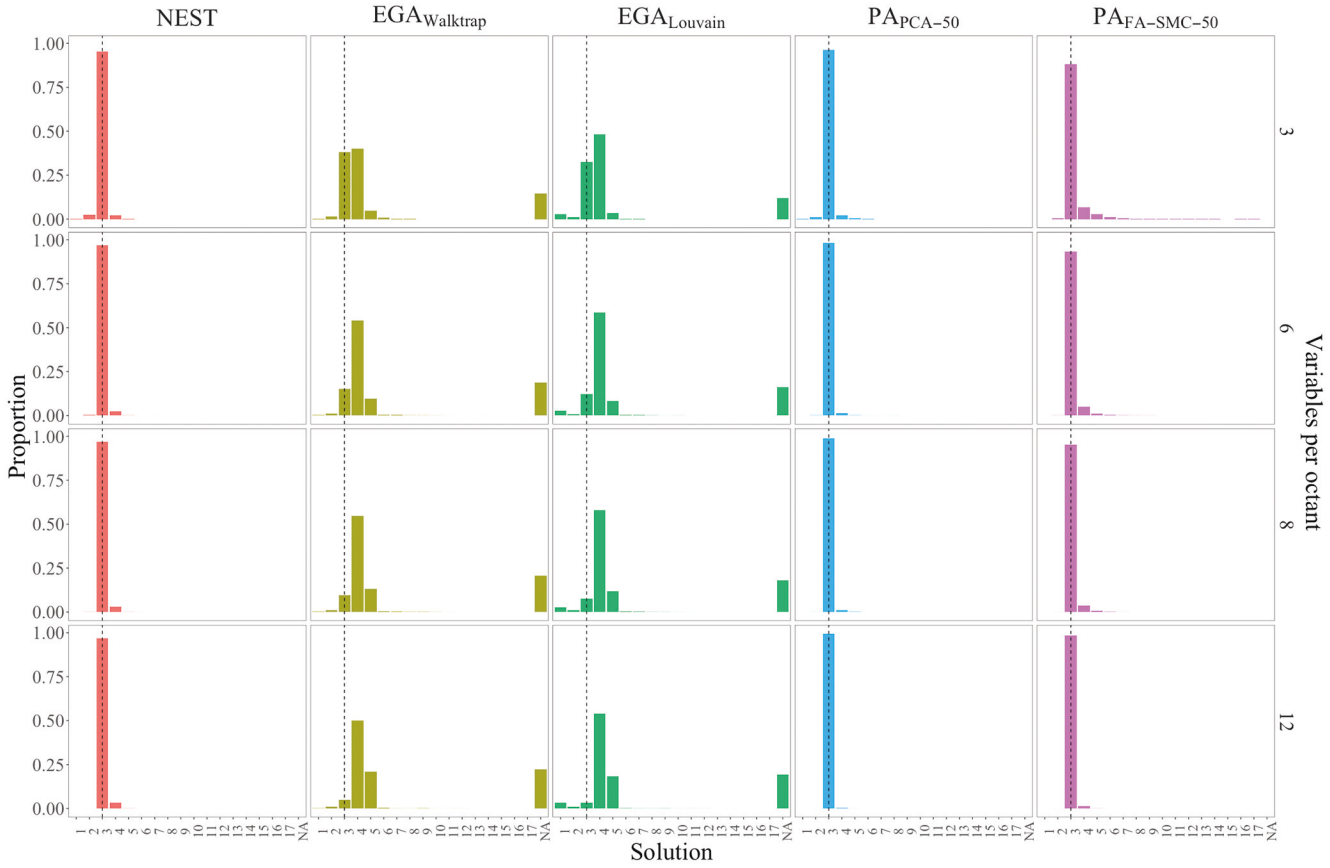
Table 4
Outcomes per Method in Simulation 2 of Study 2

	Method				
Outcome	NEST	$EGA_{Walktrap}$	$EGA_{Louvain}$	$PA_{PCA\text{-}50}$	$PA_{FA\text{-}SMC\text{-}50}$
Accurate	96.5	16.9	14.0	98.1	93.8
Underestimation	0.7	1.3	3.8	0.3	0.1
Overestimation	2.8	62.7	65.8	1.5	6.1
Undefined	0	19.0	16.4	0	0

Note. The values denote percentage proportions of outcomes. $N = 28{,}800$.

Figure 7

Solutions per Method as a Function of Variables per Octant in Simulation 2



Note. The vertical dashed line in each column marks the number of factors in the simulated circumplex models (three). The x-axis is discontinuous because it lists only solutions that were returned at least once by one method during the simulation. Undefined estimates are scored above the rightmost label (NA). N = 28,800. See the online article for the color version of this figure.

applied all methods to the eight sum scores. This way, the correlations among the analyzed variables in the method were determined through the higher-order circumplex models only. Therefore, we set the optimal number of factors in the analysis of sum scores to two in accordance to the circumplex models simulated in Simulation 3.

Results

Table 7 summarizes the outcomes of all methods in Simulation 5. Results indicate that the difference between eigenvalue-based methods and EGA was greater in Simulation 5 than in Simulation 3 under the same simulated models. EGA again showed a tendency to detect more than two factors. This is in line with other results under conditions in which analyzed correlations were predominantly determined through circumplex models.

Simulation 6

Simulated Data

In similar fashion, Simulation 6 concerns the same 76,800 data sets as Simulation 4 and computed sum scores across the eight subscales in all data sets. Here, we set the optimal number of factors to three in accordance to the higher-order circumplex models from Simulation 4, which included a general factor.

Results

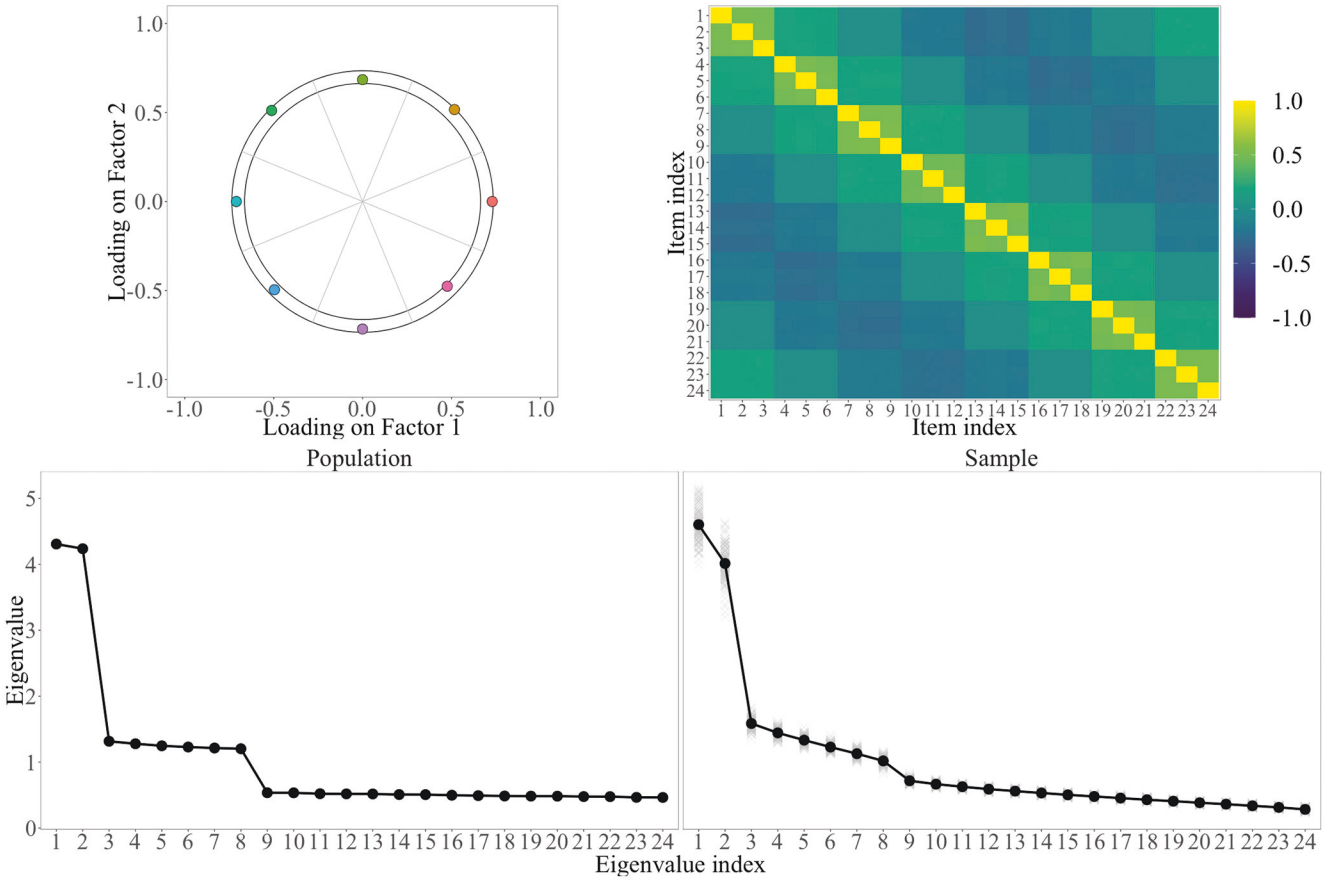
Table 8 summarizes results from Simulation 6. In contrast to Simulation 4, NEST was notably less accurate than PA; here, NEST had a stronger tendency toward underestimation and convergence failures compared to Simulation 4. Comparing Simulations 5 and 6, NEST was able to detect accurately two factors in eight sum scores (Simulation 5) but detecting three factors in eight sum scores under otherwise similar conditions was a greater challenge for NEST than for PA. EGA again was less accurate than the eigenvalue-based methods. However, while Simulation 5 showed a similar tendency toward overestimation in EGA under circumplex models as the other simulations, EGA frequently underestimated the number of factors in Simulation 6, which was not observed in the other simulations in Study 2.

Interim Discussion

Our simulations in Study 2 indicate that NEST and PA are better suited for circumplex models than EGA. The differences between the methods were greatest in Simulations 1 and 2. These

Figure 8

Illustration Circumplex Models as Simulated in Simulation 3



Note. The upper left panel denotes the loadings of the first-order factors on the two higher-order factors in the circumplex model on subscale-level. Subscales in the illustrated model meet the equal-spacing constraint. We sampled three variables per subscale with a targeted communality from U ($.70^2 - 0.05$, $.70^2 + 0.05$). The targeted communalities of the subscales in the higher-order circumplex model were also sampled from U ($.70^2 - 0.05$, $.70^2 + 0.05$). The upper right panel illustrates the correlations among the 24 variables. The bottom half of the figure illustrates the eigenvalues of the model-implied correlation matrix on the left and the distribution of eigenvalues in 150 samples under this model with $N = 250$; the gray crosses denote individual eigenvalues and the black dots connected by the black line denote the sampling mean at each index. See the online article for the color version of this figure.

results highlight how factors which were indicated in the circumplex pattern did not correspond to nonoverlapping communities that were detected by EGA. Simulations 3 and 4 imply a similar finding for combinations of independent-cluster models and circumplex models: EGA did not favor the number of factors from the circumplex model when its influence on the analyzed data sets was strongest while NEST and PA did. The gap between the methods was smaller in Simulation 5 and 6; however, Figures 5 and 7 indicate that this may have been a consequence of the reduced number of analyzed variables in Simulations 5 and 6 (always eight) compared to Simulations 1 and 2 (24 at minimum). Our simulations of circumplex models hence add to our findings from Study 1 that EGA is less compatible with substantial cross-loadings than with independent-cluster models.

Table 5

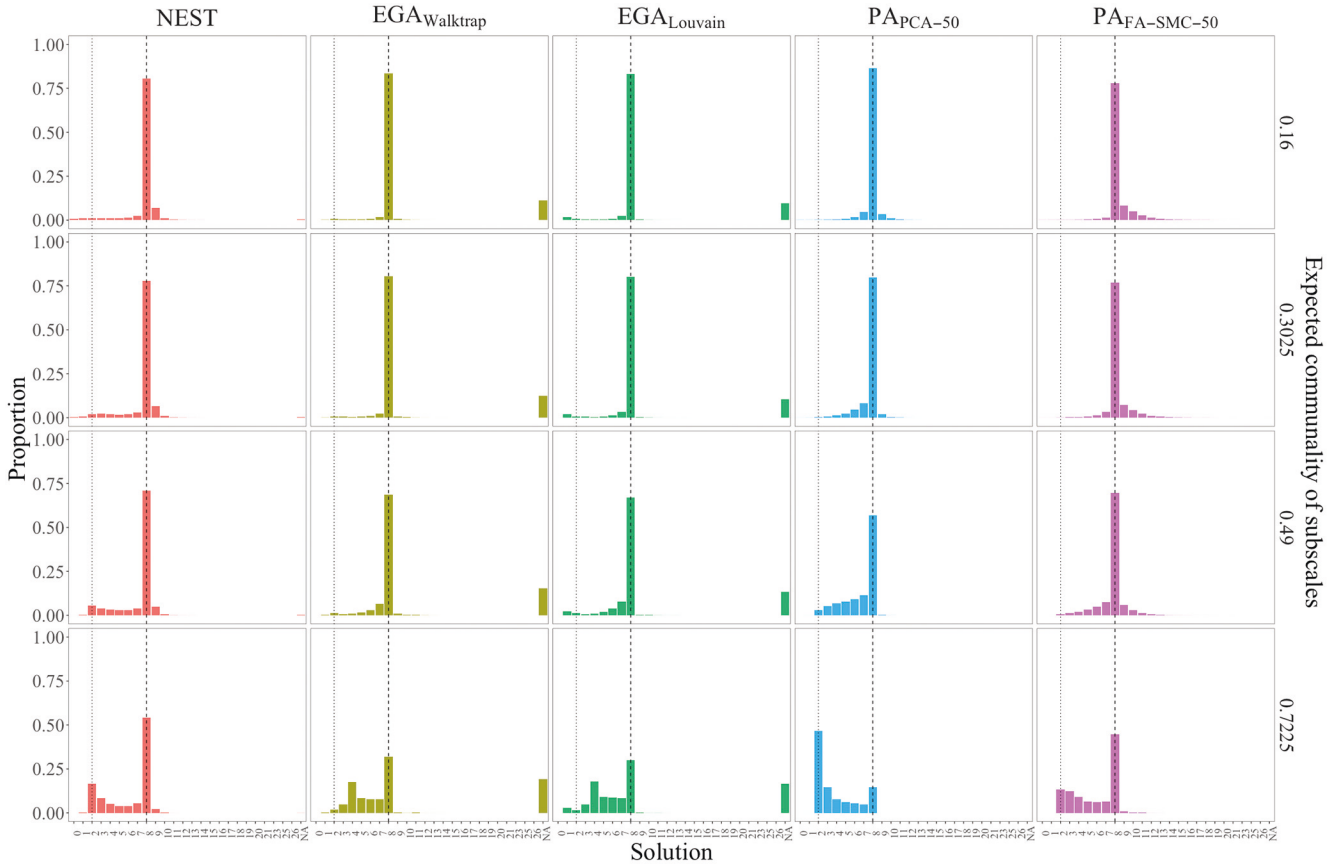
Outcomes per Method in Simulation 3 of Study 2

	Method				
Outcome	NEST	$EGA_{Walktrap}$	$EGA_{Louvain}$	PA_{PCA-50}	$PA_{FA-SMC-50}$
Accurate	70.9	66.2	65.0	59.5	67.3
Underestimation	22.7	18.2	22.0	38.5	20.7
Overestimation	6.2	1.1	0.5	2.0	12.1
Undefined	0.3	14.5	12.5	0	0

Note. The values denote percentage proportions of outcomes. $N = 76{,}800$.

Figure 9

Solutions per Method as a Function of Communalities of Subscales in Simulation 3



Note. The vertical dashed line in each column marks the number of first-order factors in the factor models on item-level (eight) and the vertical dotted line indicates the number of higher-order factors in the circumplex models on subscale-level (two). The x-axis is discontinuous because it lists only solutions that were returned at least once by one method during the simulation. Undefined estimates are scored above the rightmost label (NA). N = 76,800. See the online article for the color version of this figure.

It may be argued that Study 2 simulated problems that neither NEST nor EGA had been designed to solve in the first place, considering that methods to determine the number of factors are not commonly discussed in literature on circumplex models. Still, Study 2 shows that EGA detected nonoverlapping communities that did not correspond well to factors in circumplex models. This is a fundamental problem to EGA because it builds on the assumption that communities correspond to factors in general. In turn, Study 2 shows that EGA requires the additional assumption that data sets do not reflect circumplex models as EGA likely does not detect factors otherwise. More generally, both of our studies imply that assumptions concerning the presence of cross-loadings in all variables are needed in EGA.

Empirical Examples

Given that we designed Study 2 in an attempt to improve on ecological validity, we applied the same implementations of NEST, EGA, and PA to empirical data sets that had been discussed to reflect circumplex models. This way, we intended to explore if the differences between NEST and EGA from our simulations can also be observed outside of simulations. All reported empirical examples included negative correlations among some variables in accordance to the simulated circumplex models. In all examples, we estimated Pearson correlations among variables and passed the resulting matrices to the tested implementations.

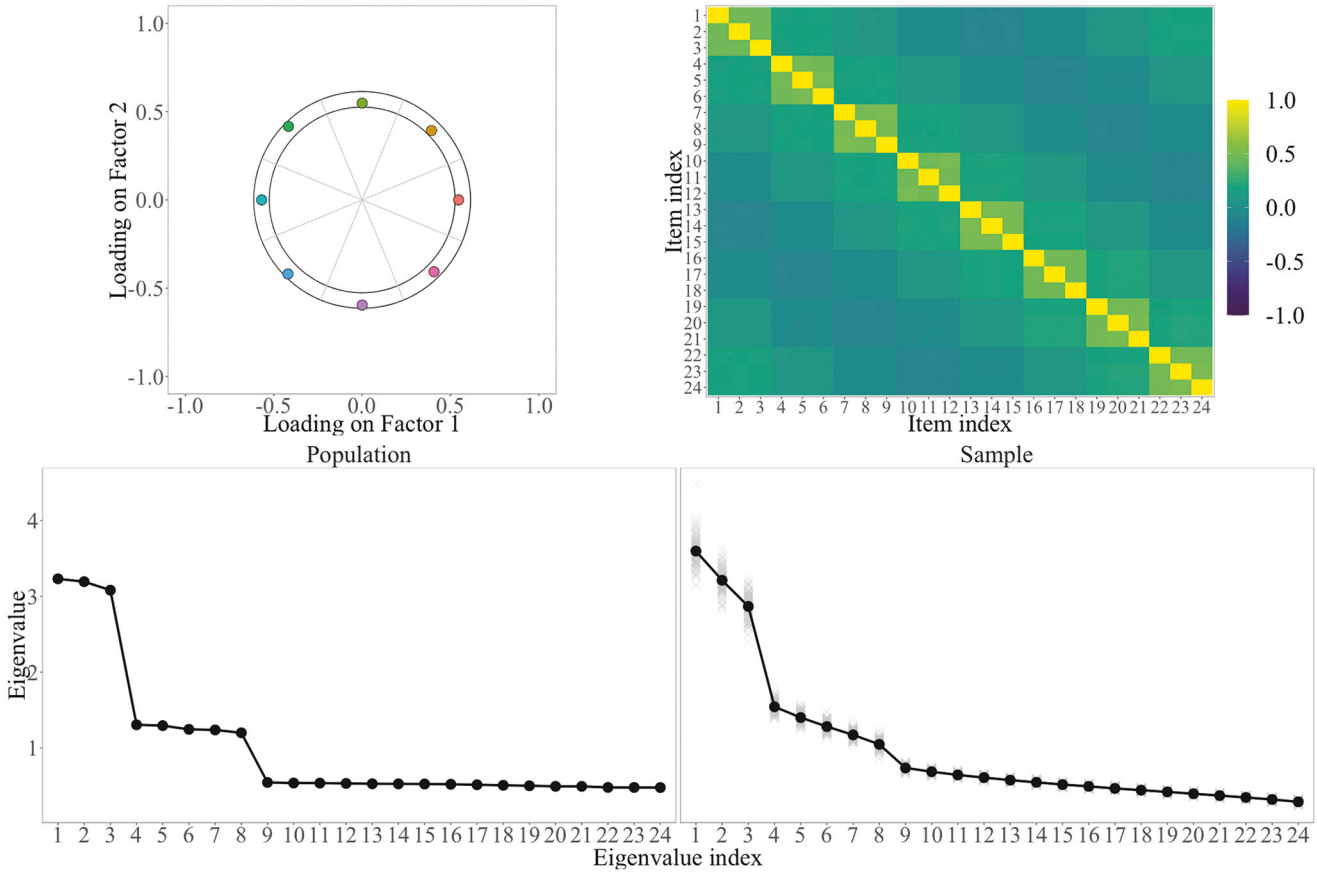
Circumplex on Item-Level

Browne (1992) developed methods to assess the fit of circumplex models and reported a correlation matrix (p. 484) as an example of correlations that show the characteristic circumplex pattern. They attributed this correlation matrix to unpublished work by William Revelle of Northwestern University (p. 483). The matrix lists correlations among 12 items from a mood questionnaire answered by 472 participants. Browne further reported loadings in a factor model of the correlations with three factors, indicating two circumplex factors and one general factor. Hence, the data set corresponds well with Simulations 1 and 2. Figure 12 provides the scree plot of eigenvalues from the correlation matrix.

NEST, PA$_{PCA-50}$, and EGA$_{Louvain}$ detected three factors, EGA$_{Walktrap}$ detected four, and PA$_{FA-SMC-50}$ detected five. Given that the fourth and subsequent factors likely explain considerably

Figure 10

Illustration Circumplex Models as Simulated in Simulation 4



Note. This figure is the pendant to Figure 9 with a general factor in the higher-order circumplex model. All conditions were the same as illustrated in Figure 9 with the exception of the general factor as evident by the addition of a third large eigenvalues compared to Figure 9. The two higher-order circumplex factors accounted for 4/6 of the communalities in all subscales. See the online article for the color version of this figure.

less variance than the preceding factors as indicated by the scree plot (see Figure 12), the solutions by $EGA_{Walktrap}$ and $PA_{FA\text{-}SMC\text{-}50}$ may be considered too liberal. This matches the bias toward overestimation from Simulation 2 in $EGA_{Walktrap}$, albeit not in $EGA_{Louvain}$.

Circumplex on Subscale-Level

Di Blas et al. (2012) developed the Interpersonal Behavior Questionnaire for Children with the aim of assessing the constructs of the interpersonal circumplex model (p. 421). It consists of 48 3-point Likert-scale items with six items each indicating a different construct corresponding to one octant of the circumplex model. Earlier, the authors had published a data set from this questionnaire with 276 complete cases as a toy data set for a software package on circumplex models (Grassi et al., 2010). While the authors expressed preference for a two-factor solution for the 48 items (Di Blas et al., 2012) from a substantive point of view, the division of items into eight subscales indicating distinct constructs on a circumplex implies ambiguity concerning the optimal number of factors on item-level. Figure 13 provides the scree plot of the

Table 6

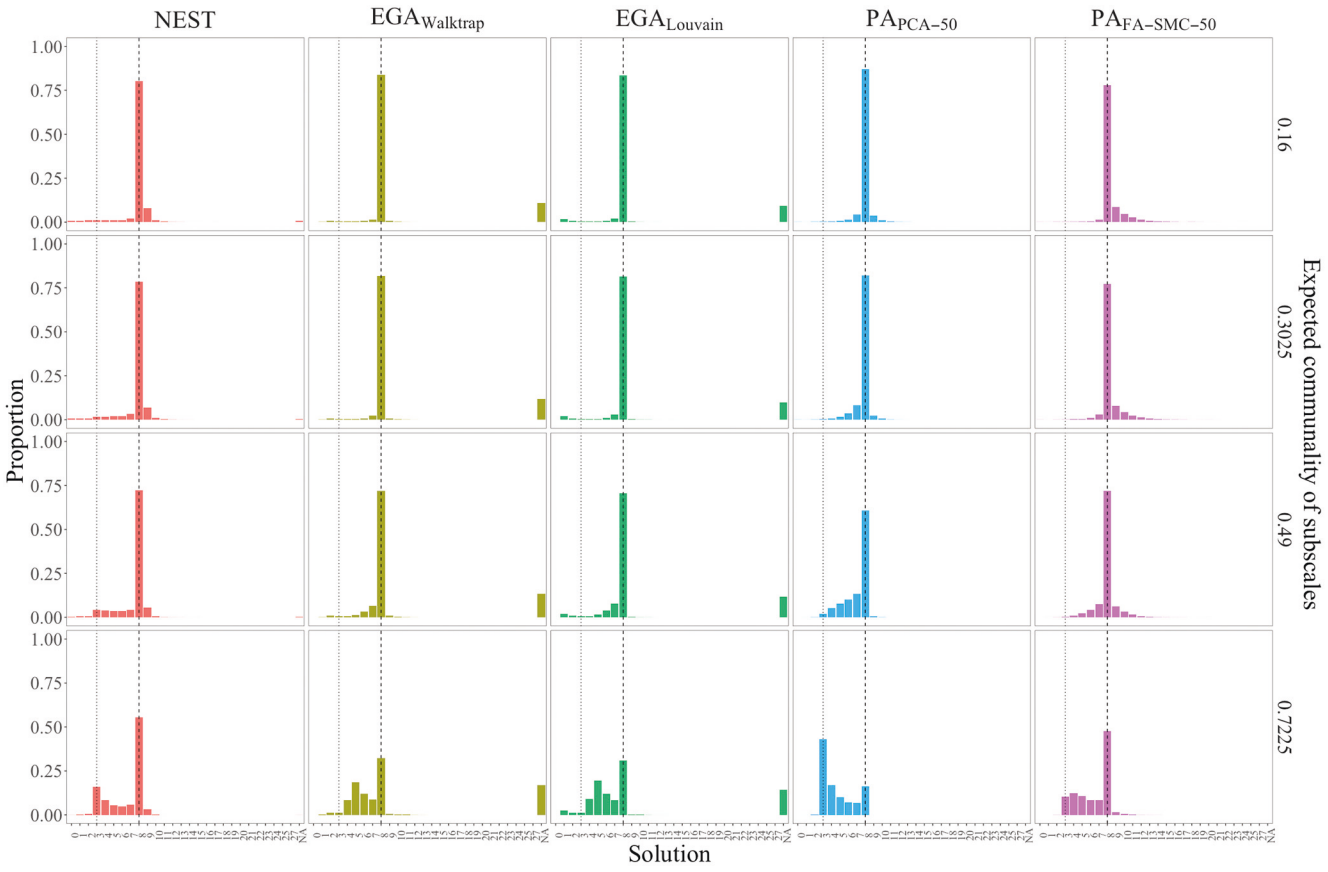
Outcomes per Method in Simulation 4 of Study 2

	Method				
Outcome	NEST	$EGA_{Walktrap}$	$EGA_{Louvain}$	$PA_{PCA\text{-}50}$	$PA_{FA\text{-}SMC\text{-}50}$
Accurate	71.6	67.3	66.7	61.5	68.6
Underestimation	21.2	18.3	21.6	36.1	18.3
Overestimation	6.9	1.1	0.5	2.3	13.0
Undefined	0.3	13.2	11.3	0	0

Note. The values denote percentage proportions of outcomes. $N = 76{,}800$.

Figure 11

Solutions per Method as a Function of Communalities of Subscales in Simulation 4



Note. The vertical dashed line in each column marks the number of first-order factors in the factor models on item-level (eight) and the vertical dotted line indicates the number of higher-order factors in the circumplex models on subscale-level (three). The x-axis is discontinuous because it lists only solutions that were returned at least once by one method during the simulation. Undefined estimates are scored above the rightmost label (NA). N = 76,800. See the online article for the color version of this figure.

correlations among items from the published data set. This data set corresponds well to Simulations 3 and 4 of Study 2.

PA$_{PCA-50}$ detected six factors, EGA$_{Louvain}$ detected seven, and NEST, EGA$_{Walktrap}$ as well as PA$_{FA-SMC-50}$ detected nine. In light of the ambiguity concerning the optimal number of factors evident in Figure 13, we refrain from labeling these solutions. However, the results correspond to the finding of Study 2 that NEST and EGA yield comparable solutions when items load on first-order factors in an independent-cluster model, indicating subscales that correlate according to a circumplex model.

Circumplex on Subscale-Aggregations

Gaines et al. (1997) reported correlations among subscales from the Interpersonal Adjective Scale (Wiggins et al., 1988) in 401 participants. This instrument contains 64 9-point Likert-scale items indicating eight subscales with eight items each. Gaines et al. (1997) computed average scores for each subscale for all participants to then compute correlations among the eight subscales. The loadings of these subscales on two factors showed a circumplex pattern. This data set provides an empirical example to our Simulations 5 and 6 from

Table 7

Outcomes per Method in Simulation 5 of Study 2

	Method				
Outcome	NEST	EGA$_{Walktrap}$	EGA$_{Louvain}$	PA$_{PCA-50}$	PA$_{FA-SMC-50}$
Accurate	89.7	67.3	62.2	92.7	85.8
Underestimation	9.2	0.9	5.9	2.5	2.0
Overestimation	0.7	10.7	15.9	4.9	12.2
Undefined	0.4	21.1	15.9	0	0

Note. The values denote percentage proportions of outcomes. N = 76,800.

Table 8
Outcomes per Method in Simulation 6 of Study 2

Outcome	Method				
	NEST	$EGA_{Walktrap}$	$EGA_{Louvain}$	PA_{PCA-50}	$PA_{FA-SMC-50}$
Accurate	72.1	42.9	51.9	86.8	82.5
Underestimation	21.9	28.3	22.9	9.2	6.5
Overestimation	0.1	3.0	5.9	4.1	11.0
Undefined	5.8	25.9	19.2	0	0

Note. The values denote percentage proportions of outcomes. $N = 76{,}800$.

Study 2. Figure 14 provides the scree plot of correlations among the eight subscales.

PA_{PCA-50} detected two factors, $EGA_{Walktrap}$ and $EGA_{Louvain}$ detected four, and $PA_{FA-SMC-50}$ detected five. Similar to the example of Browne (1992), the scree plot indicates that two factors are a reasonable solution and that higher numbers of factors may be considered overly liberal. The solutions from EGA correspond to the tendency toward overestimation in Simulation 5. NEST rejected the null-hypothesis that two factors are sufficient for the data set but ultimately failed to converge on a solution as no suitable factor model with two factors could be estimated to provide surrogate data sets.

We also applied all methods to reported correlations among the same subscales in another sample ($N = 2{,}988$; Wiggins, 1995 as cited by Gurtman & Pincus, 2000) and observed exactly the same outcomes for all methods as in the data set from Gaines et al. (1997). These results indicate that solutions in EGA are indeed questionable for subscales that indicate a circumplex model. Furthermore, they indicate frequent convergence failures in NEST for such data sets and revealed questionable rejections of two-factor solutions.

Interim Discussion

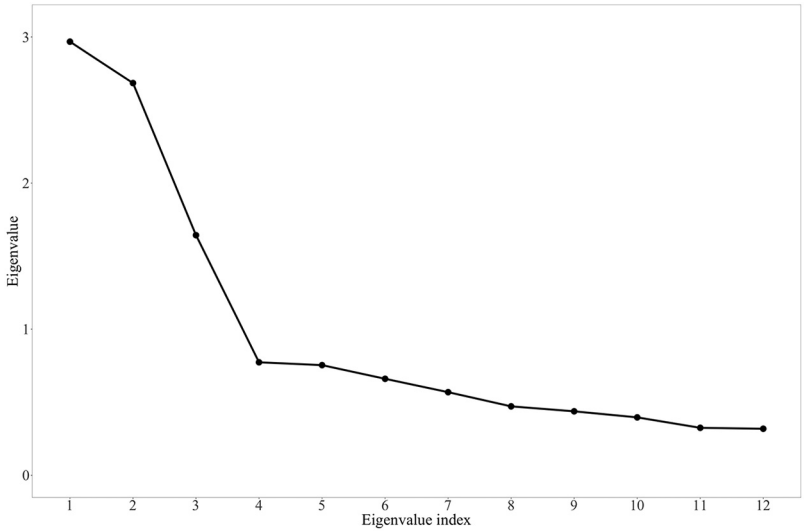
The empirical examples show solutions by EGA to disagree with solutions by other methods and also with the solutions that had been preferred by the authors that provided the analyzed data sets. This observation corresponds well to our simulations. The examples also hint at biases toward overestimation in PA_{FA} and convergence failures in NEST; these issues occurred in the simulations as well, albeit not as prominently.

However, the lack of ground truth in empirical examples in general is a limitation to our analyses. Dismissing solutions when they diverge from the solutions that authors preferred may not be reasonable in the context of circumplex models because the optimal number of factors had not been formally assessed to begin with. Furthermore, dismissing solutions that included more factors than the scree plot marked emergent may be considered superficial as the substantive nature of the factors in these solutions should be assessed in a thorough FA.

General Discussion

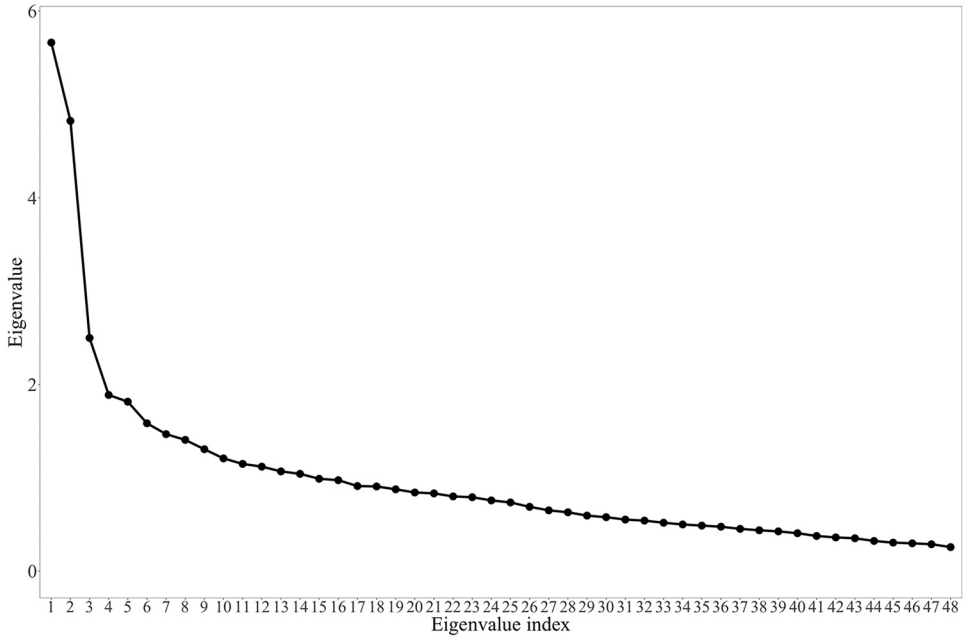
Overall, our studies highlight the need for continuous review of proposed methods to determine the optimal number of factors. On the one hand, the observed accuracy of NEST and EGA when variables loaded primarily on one factor each is in line with Achim (2017) and Golino et al. (2020). On the other hand, NEST, EGA, and PA all performed less accurately under increased item complexity (see Table 1). In Study 2, we

Figure 12
Scree Plot of Browne (1992)



Note. Browne (1992) attributed the data to unpublished work by William Revelle.

Figure 13
Scree Plot of Di Blas et al. (2012)



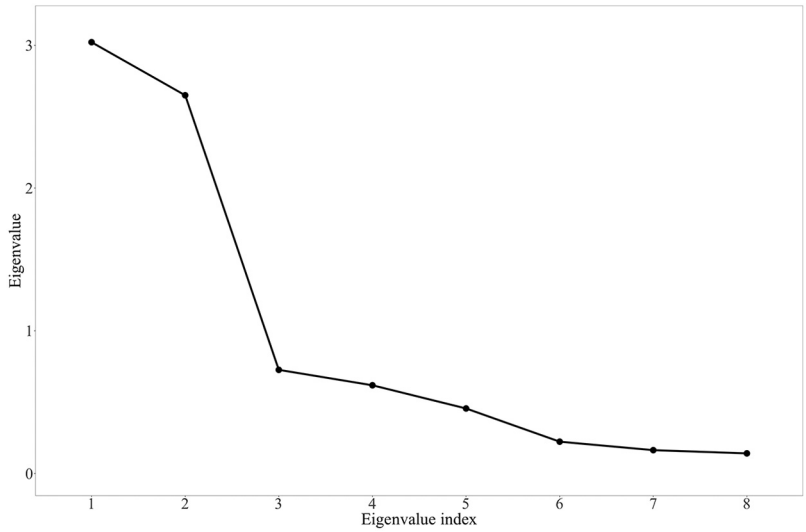
Note. This data set is available in the supplementary material of Grassi et al. (2010).

simulated circumplex models—implemented as factor models—which magnified differences between the methods regarding sensitivity to cross-loadings. NEST and PA were generally able to identify the number of factors in circumplex models whereas EGA appeared fundamentally incompatible with circumplex models. In general, our studies demonstrate that favorable performance of a method under one set of conditions does not warrant favorable performance under another set of conditions.

Implications for EGA and Network Psychometrics

EGA assumes correspondence between communities of nodes in network models and common factors in factor models. This is a prominent proposition in literature on network psychometrics (Cramer et al., 2012; Epskamp et al., 2012; Epskamp et al., 2018; van der Maas et al., 2006). The poor accuracy of EGA in the presence of cross-loadings adds nuance to this proposition. A fundamental difference between factors and nonoverlapping communities was

Figure 14
Scree Plot of Gaines et al. (1997)



best shown in Simulations 1 and 2 from Study 2: In these simulations, the simulated models unambiguously defined a set of factors and the communities detected by EGA did not correspond to these factors. The correspondence between nonoverlapping communities and factors therefore should be qualified. Independent-cluster models are a class of models in which the correspondence appears to hold; complex factor models in which items can be mapped anywhere between factors are classes of models in which it does not appear to hold.

We do not discuss the correspondence between circumplex models and network models in general; the range of all potential agreements and disagreements between circumplex models and network models is beyond the scope of our studies. Our analyses highlight one aspect of this discussion in particular, namely that solutions by EGA did not match the dimensionality in simulated factor models with substantial cross-loadings in most variables. Also, the correspondence between overlapping communities and factors (Blanken et al., 2018) remains an open question as our results only apply to nonoverlapping communities.

In its currents state, EGA requires assumptions regarding the presence of cross-loadings in analyzed data sets. This needs to be considered when EGA is employed as an alternative to EFA because assumptions on cross-loadings are not inherent to EFA: Associations between specific variables and specific factors are neither known prior to nor assumed by EFA. If such associations are known or can be assumed, CFA may be preferable to EFA (Achim, 2020) and EGA altogether.

Our studies outline avenues for further testing and development of EGA. First and foremost, further research is necessary to discuss under which conditions EGA solutions correspond to factors and under which they do not. Future studies may also discuss the substantive nature of communities when they do not correspond to factors. On a technical note, Christensen et al. (2020) discussed that overlapping communities may be implemented to account for cross-loadings. Our simulations provide data that supports this idea. Furthermore, Christensen and Golino (2021) suggested estimating the stability of EGA solutions across different realizations of analyzed variables through bootstrapping. It can be expected that problems of EGA also apply to solutions for bootstrapped samples but our results do not rule out that ill-specified solutions may be spotted through variance among EGA solutions in bootstrapped samples. Finally, the frequent convergence failures in $EGA_{Walktrap}$ and $EGA_{Louvain}$ in all of our simulations call for further research into suitable default settings for hyperparameters governing the estimation of network models and for community-detection algorithms.

Implications for NEST

In our simulations, NEST mostly performed more accurately than PA_{PCA} and about as accurately as PA_{FA}. Our results generally support the consideration of NEST in studies on methods to determine the number of factors: as subject to critical reviews, subject to further development, or as benchmark for competing methods. The supplementary material of Achim (2017) includes implementations in R—which we used in the present work—in Matlab, and in Microsoft Excel. Furthermore, the supplementary material of Achim (2020) includes an SPSS macro for NEST.

Some considerations concerning the performance of NEST in our simulations deserve discussion. First, our endorsement of NEST is at odds with other simulation studies that have compared revised parallel analysis (RPA; Green et al., 2012)—a method highly similar to NEST—to traditional PA (Auerswald & Moshagen, 2019; Lim & Jahng, 2019) and even to EGA (Cosemans et al., 2021). All three references concluded that RPA should not be preferred. The reason for the discrepancy to our results is that these studies simulated slight misfit between true population correlation matrices and underlying factor models, accounting for the fact that factor models at best approximate the data-generating process in applications on empirical data. In short, misfit was simulated through the addition of minor sources of correlations (i.e., "minor factors") to simulated factor models. RPA was reported to overestimate the number of factors from the underlying model in the presence of additional minor factors (Auerswald & Moshagen, 2019; Cosemans et al., 2021; Lim & Jahng, 2019). However, as Achim (2021) demonstrated, additional sources of correlation to simulate population misfit manifest in emergent eigenvalues that likely had been detected by RPA. We agree with Achim that detecting more factors under such conditions does not necessarily qualify as overestimation if there are indeed more sources of correlation than factors in the simulated factor model: Substantive interpretation of factors corresponding to eigenvalues cannot be applied in synthetic data sets to isolate sources of correlation that qualify as major factors rather than combinations of minor factors (Achim, 2021). Hence, we did not simulate population misfit to avoid ambiguity concerning which detectable sources of correlation qualified as factors and which did not. Consequently, our results are in line with simulation studies that attest high accuracy for RPA or NEST (Achim, 2017; Green et al., 2012; Green et al., 2015; Green et al., 2016). However, we agree with the essence of the more critical reports that factors should not be retained when they suit no theoretically sound interpretation and emphasize that NEST does not imply substantive meaningfulness of detected factors. Afterall, it should be noted that NEST rejected two-factor solutions in both empirical data sets of the Interpersonal Adjective Scale (Wiggins et al., 1988) for which two factors had been preferred in literature (Gaines et al., 1997).

While NEST was overall more accurate than PA_{PCA}, it was not consistently more accurate than PA_{FA}. Appendices A and B provide detailed comparisons of NEST to all applied variants of PA_{FA}. It is worth noting that no variant of PA_{FA} outperformed NEST throughout all simulations. Hence, advantages of PA_{FA} over NEST depended on how well simulated conditions suited the respective variant of PA_{FA}. However, discussing preference for methods should not constrain itself to accuracy but also reflect on the methods' grounding on statistical theory. Comparing NEST to PA_{FA}—or PA in general—NEST improves on a widely recognized flaw in PA: PA tests every eigenvalue from observed correlations to reference eigenvalues from surrogate data sets under null-models. The resulting sampling distribution of eigenvalues hence is not conditional on the factors that are already retained. This implies that PA provides an adequate sampling distribution for the largest eigenvalue—which has no corresponding preceding factors—but not for other eigenvalues (Achim, 2017; Braeken & van Assen,

2017; Golino et al., 2020; Green et al., 2012; Ruscio & Roche, 2012; Saccenti & Timmerman, 2017; Turner, 1998). In NEST, the threshold for all tested eigenvalues derives from a sampling distribution of eigenvalues conditional on preceding factors. Simply put, the threshold applied in PA_{FA} happened to separate the smallest eigenvalues corresponding to a factor and the largest eigenvalue not corresponding to a factor in the factor models that we simulated. This does not imply that PA_{FA} can be expected to perform as accurate under different conditions as it hardly adapts to the data sets which it is applied to.

Future research may provide a more detailed account of convergence failures in NEST and how to prevent them. One problem that may be tackled in further development is the handling of non-positive-semidefinite correlation matrices; these occurred in Study 1 and caused convergence failures in NEST in 99.7% of the affected data sets.

General Limitations

A general limitation to simulation studies is that judgment of methods does not necessarily align with criteria for sound factor retention in applied research. In applied research, the meaningfulness of a factor and the individual contribution of each factor to a model depend on the nature of the observed variables. Because all generated data sets in our simulations resulted from mere random number generation, the factors in our simulated factor models did not correspond to any domain-specific theory. Simulation studies such as the present work may reveal biases toward ill-specified solutions by inaccurate methods; accurate methods therefore may provide safeguard against statistical biases. However, accurate methods do not guarantee that all detected factors bear substantive meaningfulness nor that no meaningful factor is missed.

Conclusion

Overall, the present studies expand on reports on NEST and EGA and generated new insights into their performance in the presence of cross-loadings. NEST was sensitive to cross-loadings but able to detect accurately the number of factors in circumplex models. EGA, on the other hand, was even more sensitive to cross-loadings and failed to detect accurately the number of factors in circumplex models in many instances. Our results indicate that the number of nonoverlapping communities in network models does not correspond to factors when variables show substantial cross-loadings. All in all, our simulations provide further support for NEST and question the theoretical justification of EGA as a method to determine the optimal number of factors in exploratory analyses.

References

Achim, A. (2017). Testing the number of required dimensions in exploratory factor analysis. *The Quantitative Methods for Psychology*, *13*(1), 64–74. https://doi.org/10.20982/tqmp.13.1.p064

Achim, A. (2020). Esprit et enjeux de l'analyse factorielle exploratoire [Spirit and issues of exploratory factor analysis]. *The Quantitative Methods for Psychology*, *16*(4), 213–247. https://doi.org/10.20982/tqmp.16.4.p213

Achim, A. (2021). Determining the number of factors using parallel analysis and its recent variants: Comment on Lim and Jahng (2019). *Psychological Methods*, *26*(1), 69–73. https://doi.org/10.1037/met0000269

Acton, G. S., & Revelle, W. (2002). Interpersonal personality measures show circumplex structure based on new psychometric criteria. *Journal of Personality Assessment*, *79*(3), 446–471. https://doi.org/10.1207/S15327752JPA7903_04

Acton, G. S., & Revelle, W. (2004). Evaluation of ten psychometric criteria for circumplex structure. *Methods of Psychological Research Online*, *9*(1), 1–27.

Alden, L. E., Wiggins, J. S., & Pincus, A. L. (1990). Construction of circumplex scales for the Inventory of Interpersonal Problems. *Journal of Personality Assessment*, *55*(3-4), 521–536. https://doi.org/10.1080/00223891.1990.9674088

Auerswald, M., & Moshagen, M. (2019). How to determine the number of factors to retain in exploratory factor analysis: A comparison of extraction methods under realistic conditions. *Psychological Methods*, *24*(4), 468–491. https://doi.org/10.1037/met0000200

Blanken, T. F., Deserno, M. K., Dalege, J., Borsboom, D., Blanken, P., Kerkhof, G. A., & Cramer, A. O. (2018). The role of stabilizing and communicating symptoms given overlapping communities in psychopathology networks. *Scientific Reports*, *8*, 5854. https://doi.org/10.1038/s41598-018-24224-2

Blondel, V. D., Guillaume, J. L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics*, *2008*(10), P10008. https://doi.org/10.1088/1742-5468/2008/10/P10008

Borsboom, D. (2017). A network theory of mental disorders. *World Psychiatry*, *16*(1), 5–13. https://doi.org/10.1002/wps.20375

Borsboom, D., & Cramer, A. O. (2013). Network analysis: An integrative approach to the structure of psychopathology. *Annual Review of Clinical Psychology*, *9*(1), 91–121. https://doi.org/10.1146/annurev-clinpsy-050212-185608

Boudreaux, M. J., Ozer, D. J., Oltmanns, T. F., & Wright, A. G. C. (2018). Development and validation of the circumplex scales of interpersonal problems. *Psychological Assessment*, *30*(5), 594–609. https://doi.org/10.1037/pas0000505

Braeken, J., & van Assen, M. A. L. M. (2017). An empirical Kaiser criterion. *Psychological Methods*, *22*(3), 450–466. https://doi.org/10.1037/met0000074

Browne, M. W. (1992). Circumplex models for correlation matrices. *Psychometrika*, *57*(4), 469–497. https://doi.org/10.1007/BF02294416

Christensen, A. P., Garrido, L. E., & Golino, H. (2020). *Comparing community detection algorithms in psychological data: A Monte Carlo simulation.* arXiv. https://doi.org/10.31234/osf.io/hz89e

Christensen, A. P., & Golino, H. (2021). Estimating the stability of psychological dimensions via bootstrap exploratory graph analysis: A Monte Carlo simulation and tutorial. *Psych*, *3*(3), 479–500. https://doi.org/10.3390/psych3030032

Conway, J. M., & Huffcutt, A. I. (2003). A review and evaluation of exploratory factor analysis practices in organizational research. *Organizational Research Methods*, *6*(2), 147–168. https://doi.org/10.1177/1094428103251541

Cosemans, T., Rosseel, Y., & Gelper, S. (2021). Exploratory graph analysis for factor retention: Simulation results for continuous and binary data. *Educational and Psychological Measurement*. Advance online publication. https://doi.org/10.1177/00131644211059089

Cramer, A. O., van der Sluis, S., Noordhof, A., Wichers, M., Geschwind, N., Aggen, S. H., Kendler, K. S., & Borsboom, D. (2012). Dimensions of normal personality as networks in search of equilibrium: You can't like parties if you don't like people. *European Journal of Personality*, *26*(4), 414–431. https://doi.org/10.1002/per.1866

Cramer, A. O., Waldorp, L. J., van der Maas, H. L., & Borsboom, D. (2010). Comorbidity: A network perspective. *Behavioral and Brain Sciences*, *33*(2-3), 137–150. https://doi.org/10.1017/S0140525X09991567

Di Blas, L., Grassi, M., Luccio, R., & Momentè, S. (2012). Assessing the interpersonal circumplex model in late childhood: The interpersonal

behavior questionnaire for children. *Assessment*, *19*(4), 421–441. https://doi.org/10.1177/1073191111401172

Epskamp, S., Cramer, A. O., Waldorp, L. J., Schmittmann, V. D., & Borsboom, D. (2012). qgraph: Network visualizations of relationships in psychometric data. *Journal of Statistical Software*, *48*(4), 1–18. https://doi.org/10.18637/jss.v048.i04

Epskamp, S., Maris, G. K., Waldorp, L. J., & Borsboom, D. (2018). *Network psychometrics.* arXiv. https://doi.org/10.1002/9781118489772.ch30

Epskamp, S., Rhemtulla, M., & Borsboom, D. (2017). Generalized network psychometrics: Combining network and latent variable models. *Psychometrika*, *82*(4), 904–927. https://doi.org/10.1007/s11336-017-9557-x

Epskamp, S., Waldorp, L. J., Mõttus, R., & Borsboom, D. (2017). *Discovering psychological dynamics: The gaussian graphical model in cross-sectional and time-series data.* arXiv. https://arxiv.org/abs/1609.04156

Fabrigar, L. R., Visser, P. S., & Browne, M. W. (1997). Conceptual and methodological issues in testing the circumplex structure of data in personality and social psychology. *Personality and Social Psychology Review*, *1*(3), 184–203. https://doi.org/10.1207/s15327957pspr0103_1

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, *4*(3), 272–299. https://doi.org/10.1037/1082-989X.4.3.272

Fava, J. L., & Velicer, W. F. (1992). The effects of overextraction on factor and component analysis. *Multivariate Behavioral Research*, *27*(3), 387–415. https://doi.org/10.1207/s15327906mbr2703_5

Fried, E. I., & Cramer, A. O. J. (2017). Moving forward: Challenges and directions for psychopathological network theory and methodology. *Perspectives on Psychological Science*, *12*(6), 999–1020. https://doi.org/10.1177/1745691617705892

Friedman, J., Hastie, T., & Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, *9*(3), 432–441. https://doi.org/10.1093/biostatistics/kxm045

Gaines, S. O., Jr., Panter, A. T., Lyde, M. D., Steers, W. N., Rusbult, C. E., Cox, C. L., & Wexler, M. O. (1997). Evaluating the circumplexity of interpersonal traits and the manifestation of interpersonal traits in interpersonal trust. *Journal of Personality and Social Psychology*, *73*(3), 610–623. https://doi.org/10.1037/0022-3514.73.3.610

Genz, A., Bretz, F., Miwa, T., Mi, X., Leisch, F., Scheipl, F., & Hothorn, T. (2021). *mvtnorm: Multivariate normal and t distributions.* http://cran.r-project.org/package=mvtnorm

Golino, H. F., & Epskamp, S. (2017). Exploratory graph analysis: A new approach for estimating the number of dimensions in psychological research. *PLoS ONE*, *12*(6), e0174035. https://doi.org/10.1371/journal.pone.0174035

Golino, H., Shi, D., Christensen, A. P., Garrido, L. E., Nieto, M. D., Sadana, R., Thiyagarajan, J. A., & Martinez-Molina, A. (2020). Investigating the performance of exploratory graph analysis and traditional techniques to identify the number of latent factors: A simulation and tutorial. *Psychological Methods*, *25*(3), 292–320. https://doi.org/10.1037/met0000255

Golino, H., & Christensen, A. P. (2021). *EGAnet: Exploratory graph analysis—A framework for estimating the number of dimensions in multivariate data using network psychometrics.* http://cran.r-project.org/package=EGAnet

Goretzko, D., & Bühner, M. (2020). One model to rule them all? Using machine learning algorithms to determine the number of factors in exploratory factor analysis. *Psychological Methods*, *25*(6), 776–786. https://doi.org/10.1037/met0000262

Grassi, M., Luccio, R., & Di Blas, L. (2010). CircE: An R implementation of Browne's circular stochastic process model. *Behavior Research Methods*, *42*(1), 55–73. https://doi.org/10.3758/BRM.42.1.55

Green, S. B., Levy, R., Thompson, M. S., Lu, M., & Lo, W. J. (2012). A proposed solution to the problem with using completely random data to assess the number of factors with parallel analysis. *Educational and Psychological Measurement*, *72*(3), 357–374. https://doi.org/10.1177/0013164411422252

Green, S. B., Redell, N., Thompson, M. S., & Levy, R. (2016). Accuracy of revised and traditional parallel analyses for assessing dimensionality with binary data. *Educational and Psychological Measurement*, *76*(1), 5–21. https://doi.org/10.1177/0013164415581898

Green, S. B., Thompson, M. S., Levy, R., & Lo, W. J. (2015). Type I and type II error rates and overall accuracy of the revised parallel analysis method for determining the number of factors. *Educational and Psychological Measurement*, *75*(3), 428–457. https://doi.org/10.1177/0013164414546566

Gurtman, M. B., & Pincus, A. L. (2000). Interpersonal adjective scales: Confirmation of circumplex structure from multiple perspectives. *Personality and Social Psychology Bulletin*, *26*(3), 374–384. https://doi.org/10.1177/0146167200265009

Gurtman, M. B., & Pincus, A. L. (2003). The circumplex model: Methods and research applications. In J. A. Schinka & W. F. Velicer (Eds.), *Handbook of psychology*: *Vol. 2. Research methods in psychology* (pp. 407–428). Wiley.

Guttman, L. (1954). A new approach to factor analysis: The radex. In P. F. Lazarsfeld (Ed.), *Mathematical thinking in the social sciences* (pp. 258–348). Columbia University Press.

Hallquist, M. N., Wright, A. G. C., & Molenaar, P. C. M. (2021). Problems with centrality measures in psychopathology symptom networks: Why network psychometrics cannot escape psychometric theory. *Multivariate Behavioral Research*, *56*(2), 199–223. https://doi.org/10.1080/00273171.2019.1640103

Henson, R. K., & Roberts, J. K. (2006). Use of exploratory factor analysis in published research: Common errors and some comment on improved practice. *Educational and Psychological Measurement*, *66*(3), 393–416. https://doi.org/10.1177/0013164405282485

Hopwood, C. J., Ansell, E. B., Pincus, A. L., Wright, A. G., Lukowitsky, M. R., & Roche, M. J. (2011). The circumplex structure of interpersonal sensitivities. *Journal of Personality*, *79*(4), 707–740. https://doi.org/10.1111/j.1467-6494.2011.00696.x

Hopwood, C. J., & Good, E. W. (2019). Structure and correlates of interpersonal problems and sensitivities. *Journal of Personality*, *87*(4), 843–855. https://doi.org/10.1111/jopy.12437

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*(2), 179–185. https://doi.org/10.1007/BF02289447

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*(4), 434–446. https://doi.org/10.1016/j.jml.2007.11.007

Kan, K. J., de Jonge, H., van der Maas, H. L. J., Levine, S. Z., & Epskamp, S. (2020). How to compare psychometric factor and network models. *Journal of Intelligence*, *8*(4), 35. https://doi.org/10.3390/jintelligence8040035

Keith, T. Z., Caemmerer, J. M., & Reynolds, M. R. (2016). Comparison of methods for factor extraction for cognitive test-like data: Which overfactor, which underfactor? *Intelligence*, *54*, 37–54. https://doi.org/10.1016/j.intell.2015.11.003

Lim, S., & Jahng, S. (2019). Determining the number of factors using parallel analysis and its recent variants. *Psychological Methods*, *24*(4), 452–467. https://doi.org/10.1037/met0000230

Locke, K. D. (2000). Circumplex scales of interpersonal values: Reliability, validity, and applicability to interpersonal problems and personality disorders. *Journal of Personality Assessment*, *75*(2), 249–267. https://doi.org/10.1207/S15327752JPA7502_6

Lorenzo-Seva, U., Timmerman, M. E., & Kiers, H. A. (2011). The Hull method for selecting the number of common factors. *Multivariate Behavioral Research*, *46*(2), 340–364. https://doi.org/10.1080/00273171.2011.564527

Marsman, M., Borsboom, D., Kruis, J., Epskamp, S., van Bork, R., Waldorp, L. J., Maas, H. L. J. V., & Maris, G. (2018). An introduction to network psychometrics: Relating Ising network models to item response theory models. *Multivariate Behavioral Research*, *53*(1), 15–35. https://doi.org/10.1080/00273171.2017.1379379

Neal, Z. P., & Neal, J. W. (2021). Out of bounds? The boundary specification problem for centrality in psychological networks. *Psychological Methods*. Advance online publication. https://doi.org/10.1037/met0000426

Pons, P., & Latapy, M. (2006). Computing communities in large networks using random walks. *Journal of Graph Algorithms and Applications*, *10*(2), 191–218. https://doi.org/10.7155/jgaa.00124

Preacher, K. J., Zhang, G., Kim, C., & Mels, G. (2013). Choosing the optimal number of factors in exploratory factor analysis: A model selection perspective. *Multivariate Behavioral Research*, *48*(1), 28–56. https://doi.org/10.1080/00273171.2012.710386

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.r-project.org/

Rafaeli, E., & Revelle, W. (2006). A premature consensus: Are happiness and sadness truly opposite affects? *Motivation and Emotion*, *30*(1), 1–12. https://doi.org/10.1007/s11031-006-9004-2

Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research*, *47*(5), 667–696. https://doi.org/10.1080/00273171.2012.715555

Remmington, N. A., Fabrigar, L. R., & Visser, P. S. (2000). Reexamining the circumplex model of affect. *Journal of Personality and Social Psychology*, *79*(2), 286–300. https://doi.org/10.1037/0022-3514.79.2.286

Revelle, W., & Rocklin, T. (1979). Very simple structure: An alternative procedure for estimating the optimal number of interpretable factors. *Multivariate Behavioral Research*, *14*(4), 403–414. https://doi.org/10.1207/s15327906mbr1404_2

Revelle, W. (2022). *Psych: Procedures for psychological, psychometric, and personality research*. https://cran.r-project.org/package=psych

Revelle, W. (n.d.). *An introduction to psychometric theory with applications in R*. https://personality-project.org/r/book/

Rogoza, R., Cieciuch, J., & Strus, W. (2021). A three-step procedure for analysis of circumplex models: An example of narcissism located within the circumplex of personality metatraits. *Personality and Individual Differences*, *169*, 109775. https://doi.org/10.1016/j.paid.2019.109775

Ruscio, J., & Roche, B. (2012). Determining the number of factors to retain in an exploratory factor analysis using comparison data of known factorial structure. *Psychological Assessment*, *24*(2), 282–292. https://doi.org/10.1037/a0025697

Saccenti, E., & Timmerman, M. E. (2017). Considering Horn's parallel analysis from a random matrix theory point of view. *Psychometrika*, *82*(1), 186–209. https://doi.org/10.1007/s11336-016-9515-z

Timmerman, M. E., & Lorenzo-Seva, U. (2011). Dimensionality assessment of ordered polytomous items with parallel analysis. *Psychological Methods*, *16*(2), 209–220. https://doi.org/10.1037/a0023353

Tracey, T. J. (2000). Analysis of circumplex models. In H. E. A. Tinsley & S. D. Brown (Eds.), *Handbook of applied multivariate statistics and mathematical modeling* (pp. 641–664). Academic Press. https://doi.org/10.1016/B978-012691360-6/50023-9

Turner, N. E. (1998). The effect of common variance and structure pattern on random data eigenvalues: Implications for the accuracy of parallel analysis. *Educational and Psychological Measurement*, *58*(4), 541–568. https://doi.org/10.1177/0013164498058004001

van Bork, R., Epskamp, S., Rhemtulla, M., Borsboom, D., & van der Maas, H. L. (2017). What is the p-factor of psychopathology? Some risks of general factor modeling. *Theory & Psychology*, *27*(6), 759–773. https://doi.org/10.1177/0959354317737185

van der Maas, H. L., Dolan, C. V., Grasman, R. P., Wicherts, J. M., Huizenga, H. M., & Raijmakers, M. E. (2006). A dynamical model of general intelligence: The positive manifold of intelligence by mutualism. *Psychological Review*, *113*(4), 842–861. https://doi.org/10.1037/0033-295X.113.4.842

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*(6), 1063–1070. https://doi.org/10.1037/0022-3514.54.6.1063

Widaman, K. F. (2018). On common factor and principal component representations of data: Implications for theory and for confirmatory replications. *Structural Equation Modeling*, *25*(6), 829–847. https://doi.org/10.1080/10705511.2018.1478730

Wiggins, J. S., Trapnell, P., & Phillips, N. (1988). Psychometric and geometric characteristics of the Revised Interpersonal Adjective Scales (IAS-R). *Multivariate Behavioral Research*, *23*(4), 517–530. https://doi.org/10.1207/s15327906mbr2304_8

Wright, A. G., Pincus, A. L., Conroy, D. E., & Hilsenroth, M. J. (2009). Integrating methods to optimize circumplex description and comparison of groups. *Journal of Personality Assessment*, *91*(4), 311–322. https://doi.org/10.1080/00223890902935696

Ziegler, M., & Hagemann, D. (2015). Testing the unidimensionality of items. *European Journal of Psychological Assessment*, *31*(4), 231–237. https://doi.org/10.1027/1015-5759/a000309

Zimmermann, J., & Wright, A. G. (2017). Beyond description in interpersonal construct validation: Methodological advances in the circumplex structural summary approach. *Assessment*, *24*(1), 3–23. https://doi.org/10.1177/1073191115621795

Zwick, W. R., & Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin*, *99*(3), 432–442. https://doi.org/10.1037/0033-2909.99.3.432

(*Appendices follow*)

Appendix A

Detailed Account for Parallel Analysis in Study 1

Appendix A provides a more detailed account of PA variants in the results of Study 1. The Tables A1 and A2 refer to the same data sets as Tables 1 and 2, respectively.

Table A1

Outcomes per Method in Study 1

		Method								
Outcome	Item complexity	NEST	$EGA_{Walktrap}$	$EGA_{Louvain}$	$PA_{PCA\text{-}95}$	$PA_{PCA\text{-}50}$	$PA_{FA\text{-}SMC\text{-}95}$	$PA_{FA\text{-}SMC\text{-}50}$	$PA_{FA\text{-}1F\text{-}95}$	$PA_{FA\text{-}1F\text{-}50}$
Accurate	1	84.2	79.2	79.8	71.6	75.6	84.0	83.3	80.9	85.6
	2	49.5	21.1	24.4	24.2	27.7	46.5	53.7	46.3	54.9
Underestimation	1	14.0	8.0	9.3	28.2	22.9	14.4	8.0	18.2	6.8
	2	41.2	53.6	50.0	75.8	72.0	52.9	41.7	53.2	39.0
Overestimation	1	1.5	1.2	1.4	0.2	1.5	1.6	8.7	0.9	7.7
	2	0.8	7.9	10.5	0	0.4	0.6	4.4	0.4	6.1
Undefined	1	0.2	11.6	9.5	0	0	0	0	0	0
	2	8.5	17.4	15.1	0	0	0.1	0.2	0	0

Note. The values denote percentage proportions of outcomes. $N = 307{,}200$.

Table A2

Accuracy per Method as a Function of the Number of Factors and Item Complexity

		Method								
No. of factors	Item complexity	NEST	$EGA_{Walktrap}$	$EGA_{Louvain}$	$PA_{PCA\text{-}95}$	$PA_{PCA\text{-}50}$	$PA_{FA\text{-}SMC\text{-}95}$	$PA_{FA\text{-}SMC\text{-}50}$	$PA_{FA\text{-}1F\text{-}95}$	$PA_{FA\text{-}1F\text{-}50}$
2 factors	1	96.0	90.3	90.2	97.4	97.4	96.8	92.4	83.0	92.1
	2	42.7	23.2	15.2	39.9	47.4	66.3	74.1	68.9	75.2
3 factors	1	94.2	88.9	88.9	96.3	95.7	95.2	88.8	83.8	92.4
	2	80.2	46.8	53.7	62.2	68.9	80.6	82.6	78.9	82.7
4 factors	1	91.7	87.5	87.6	94.2	94.4	93.3	86.6	84.4	92.0
	2	78.3	30.0	29.6	61.5	67.3	79.0	80.0	74.1	79.4
5 factors	1	89.4	86.4	86.5	92.2	92.7	91.1	83.3	87.1	90.6
	2	75.3	25.5	23.0	58.0	62.7	75.3	77.1	67.6	72.8

Note. The values denote percentage proportions of accurate outcomes. $N = 25{,}600$.

(*Appendices continue*)

Appendix B

Detailed Account for Parallel Analysis in Study 2

Appendix B summarizes results from the six simulations of Study 2 in Table B1, which also accounts for all PA variants.

Table B1

Outcomes per Method in All Simulations of Study 2

		Method								
Simulation	Outcome	NEST	$EGA_{Walktrap}$	$EGA_{Louvain}$	PA_{PCA-95}	PA_{PCA-50}	$PA_{FA-SMC-95}$	$PA_{FA-SMC-50}$	$PA_{FA-1F-95}$	$PA_{FA-1F-50}$
Simulation 1	Accurate	98.0	6.9	2.9	99.9	99.2	99.5	96.8	99.6	97.8
N = 9,600	Underestimation	0	0.1	2.6	0	0	0	0	0	0
	Overestimation	2.0	73.6	77.7	0.1	0.8	0.5	3.2	0.4	2.2
	Undefined	0	19.3	16.9	0	0	0	0	0	0
Simulation 2	Accurate	96.5	16.9	14.0	98.8	98.1	98.1	93.8	98.7	96.4
N = 28,800	Underestimation	0.7	1.3	3.8	0.9	0.3	0.6	0.1	0.6	0.1
	Overestimation	2.8	62.7	65.8	0.3	1.5	1.3	6.1	0.7	3.4
	Undefined	0	19.0	16.4	0	0	0	0	0	0
Simulation 3	Accurate	70.9	66.2	65.0	56.5	59.5	68.2	67.3	63.1	65.9
N = 76,800	Underestimation	22.7	18.2	22.0	43.2	38.5	27.6	20.7	36.0	30.1
	Overestimation	6.2	1.1	0.5	0.3	2.0	4.2	12.1	0.8	4.0
	Undefined	0.3	14.5	12.5	0	0	0	0	0	0
Simulation 4	Accurate	71.6	67.3	66.7	58.3	61.5	69.6	68.6	63.5	66.4
N = 76,800	Underestimation	21.2	18.3	21.6	41.2	36.1	25.8	18.3	35.6	29.6
	Overestimation	6.9	1.1	0.5	0.5	2.3	4.6	13.0	0.9	4.0
	Undefined	0.3	13.2	11.3	0	0	0	0	0	0
Simulation 5	Accurate	89.7	67.3	62.2	91.4	92.7	90.9	85.8	61.9	81.3
N = 76,800	Underestimation	9.2	0.9	5.9	8.2	2.5	7.5	2.0	37.8	2.8
	Overestimation	0.7	10.7	15.9	0.4	4.9	1.6	12.2	0.3	15.9
	Undefined	0.4	21.1	15.9	0	0	0	0	0	0
Simulation 6	Accurate	72.1	42.9	51.9	78.5	86.8	81.1	82.5	42.6	79.3
N = 76,800	Underestimation	21.9	28.3	22.9	21.4	9.2	18.0	6.5	57.1	7.9
	Overestimation	0.1	3.0	5.9	0.1	4.1	0.9	11.0	0.3	12.8
	Undefined	5.8	25.9	19.2	0	0	0	0	0	0

Note. The values denote percentage proportions of outcomes.

Received February 12, 2021
Revision received May 30, 2022
Accepted June 15, 2022 ■

Factor Retention in Ordered Categorical Variables: Benefits and Costs of Polychoric Correlations in Eigenvalue-Based Testing

Nils Brandenburg[1]

[1]Heinrich Heine University Düsseldorf

Corresponding author: Nils Brandenburg

Department of Experimental Psychology

Heinrich Heine University Düsseldorf

Universitätsstr. 1

40225 Düsseldorf, Germany

Tel.: +49 211 81-14568

Email: nils.brandenburg@hhu.de

Abstract

An essential step in exploratory factor analysis is to determine the optimal number of factors. The Next Eigenvalue Sufficiency Test (NEST; Achim, 2017) is a recent proposal to determine the number of factors based on significance tests of the statistical contributions of candidate factors indicated by eigenvalues of sample correlation matrices. Previous simulation studies that have shown NEST to recover the optimal number of factors in simulated datasets with high accuracy. However, these studies have focused on continuous variables. The present work addresses the performance of NEST for categorical data. It has been debated whether factor models for categorical variables should be applied to Pearson correlation matrices, which are known to underestimate correlations for categorical datasets, or to polychoric correlation matrices, which are known to be instable. The central research question is to what extent the problems that come with Pearson correlations and polychoric correlations deteriorate NEST for categorical variables. An implementation of NEST in which polychoric correlations are computed for categorical datasets is proposed. The proposed implementation and the original implementation of NEST in which Pearson correlations are computed are compared in a simulation study. The simulation shows that using polychoric correlations improves the performance of NEST compared to using Pearson correlations for binary variables and large sample size ($N = 500$). However, the simulation also shows that NEST performed better with Pearson correlations than with polychoric correlations for Likert-type variables with four response categories when item difficulties were homogeneous.

Keywords: Exploratory factor analysis, factor retention, Next Eigenvalue Sufficiency Test, categorical variables, polychoric correlations.

Factor Retention for Categorical Variables: Benefits and Problems of Polychoric Correlations

Factor analysis is a common method to model correlations among a set of variables as functions of a smaller number of common factors. Historically, factor analysis has played a central role in scale development as it provides assessment to what extent sets of items collectively measure a common construct (i.e., common factors; Conway & Huffcutt, 2003; Henson & Roberts, 2006; O'Leary-Kelly & Vorkuka, 1998; Ziegler & Hagemann, 2015). Technically, factor analysis is used to estimate a model of the population correlation matrix for a set of variables. For a brief introduction, let p be the number of analyzed variables and Σ be the $p \times p$ correlation matrix and let m be the number of common factors in the factor model of Σ. The linear factor model of Σ is given in the following Equation:

$$\Sigma = \Lambda\Psi\Lambda^{\mathrm{T}} + \Theta \quad (1)$$

In Equation 1, Λ is the $p \times m$ matrix that denotes regression weights (i.e., factor loadings) of common factors on the variables, Ψ is the $m \times m$ matrix of correlations among common factors, and Θ is the $p \times p$ matrix that increments $\Lambda\Psi\Lambda^{\mathrm{T}}$ by contributions of components that are dissociated from common factors in that they are unique to each variable. In the following sections, 'common factors' are referred to simply as factors and 'factor loadings' are referred to as loadings.

A special case of factor analysis is exploratory factor analysis which has been designed specifically to explore links between factors and variables when no factors can be specified on theoretical grounds (Achim, 2020; Widaman, 2018). To this end, exploratory factor analysis estimates factor models without constraining loading parameters to 0. When it cannot be specified which factors inform which variable, it is likely also unknown how many factors can be assumed to underlie the dataset to begin with. Hence, exploratory factor analysis

typically includes some formal determination of the number of factors prior to the parameter estimation for the final model (Fabrigar et al., 1999; Fava & Velicer, 1992; Goretzko et al., 2021). It is key to determine the optimal number of factors so that the factor model retains all factors of substantive importance while none of the identified factors is spurious (Auerswald & Moshagen, 2019; Braeken & van Assen, 2017; Fabrigar et al., 1999; Fava & Velicer, 1992; Henson & Roberts, 2006; Preacher et al., 2013; Schmitt, 2011). The problem of determining the optimal number of factors is referred to as the number-of-factors problem.

Consensus on how to approach the number-of-factors problem in empirical research has yet to be reached. In fact, numerous methods to determine the number of factors have been proposed in the past decade (Achim, 2017; Braeken & van Assen, 2017; Green et al., 2012; Golino & Epskamp, 2017; Goretzko & Bühner, 2020; Ruscio & Roche, 2012). One of the recent proposals is a method coined Next Eigenvalue Sufficiency Test (NEST; Achim, 2017). The objective of the present work is to contribute to the validation and the development of NEST. Previous simulation studies have shown that NEST determines the number of factors accurately compared to other methods (Achim, 2017; Brandenburg & Papenberg, 2022), but so far, evidence from simulations has been limited to continuous variables. Here, datasets with categorical variables were simulated in order to test whether NEST also determines the number of factors accurately for categorical variables. It was tested whether there is an optimal implementation of NEST for categorical variables. In the following sections, it is introduced how NEST aims to determine the optimal number of factors for correlations among analyzed variables. Then, it is outlined how categorical variables challenge NEST and how its performance can be expected to depend on the computation of correlations. Subsequently, an implementation of NEST that is tailored to categorical variables is proposed. Finally, a simulation study is reported in which the proposed implementation and the original implementation are compared for simulated categorical variables.

Next Eigenvalue Sufficiency Test

In general, NEST determines the number of factors in a dataset through examination of the eigenvalues of the sample correlation matrix. Eigenvalues of sample correlation matrices are also central to other methods to determine the number of factors, such as the eigenvalue-greater-than-1 rule (Guttman, 1954), the Scree Test (Cattell, 1966), and Parallel Analysis (PA; Horn, 1965). To understand the importance of eigenvalues in the context of factor analysis, consider the term $\Lambda\Psi\Lambda^{\mathrm{T}}$ from Equation 1. This term accounts for all common-factor related parameters; it constitutes a $p \times p$ matrix that lists model-implied pairwise correlation coefficients as off-diagonal elements and nonunique variances in variables on the main diagonal. The model-implied nonunique variance in a variable is commonly referred to as the variable's communality. Crucially, the k^{th} largest eigenvalue obtained from eigenvalue-decomposition of $\Lambda\Psi\Lambda^{\mathrm{T}}$ is equal to the increment of communality across all variables that can be attributed to the k^{th} factor. Hence, eigenvalues of $\Lambda\Psi\Lambda^{\mathrm{T}}$ indicate the amount of variance that each factor contributes to a factor model, which provides useful information concerning the number of factors to be retained.

Of course, in empirical applications, no true model $\Lambda\Psi\Lambda^{\mathrm{T}}$ that could be examined prior to factor analysis is known. Therefore, eigenvalue-based methods like NEST are commonly based on the eigenvalues of sample correlation matrices. Unlike $\Lambda\Psi\Lambda^{\mathrm{T}}$, eigenvalues of a sample correlation matrix do not indicate the exact amount of variance that factors contribute to a factor model. In fact, eigenvalues of the sample correlation matrix indicate the amount of variance that principal components contribute in principal component analysis (Widaman, 2018). The widely accepted reasoning of methods to determine the number of factors in factor analysis based on eigenvalues of sample correlation matrices can be summarized as follows: If the data-generating process is best depicted as a factor model with m factors, the m largest eigenvalues of the sample correlation matrix are expected to be sufficiently larger than all remaining eigenvalues. What separates eigenvalue-based methods from each other is how

they distinguish eigenvalues that indicate the presence of a factor from eigenvalues that do not.

NEST is an iterative testing procedure of the null hypothesis that k factors are sufficient to estimate a factor model that fits the analyzed dataset, starting with $k = 0$. The relevant test statistic that indicates the presence of at least $k + 1$ factors—as an alternative to the null hypothesis of k-factor sufficiency—is the eigenvalue at index $k + 1$. To test the null hypothesis, NEST first computes a reference model with k principal axis factors of the dataset and then simulates j surrogate datasets under the k-factor model with the same sample size and number of variables as the dataset in question. For $k = 0$, surrogate datasets are sampled from a population of independent variables. The eigenvalues at index $k + 1$ thus form a sampling distribution of eigenvalues conditional on the k-factor model fitting the analyzed dataset. NEST makes no distributional assumptions concerning the distribution of eigenvalues and ranks the tested eigenvalue among the simulated sampling distribution to test the null hypothesis. For a given α level, the null hypothesis is rejected if the rank of the tested eigenvalue is greater than $(1 - \alpha)(j + 1)$, given that higher ranks correspond to larger eigenvalues.

By updating the sampling distribution of tested eigenvalues conditional on the factors for which retention has already been confirmed, NEST differs from the well-studied PA, which also ranks eigenvalues of the analyzed dataset among eigenvalues of surrogate datasets. The main difference between NEST and PA is that PA samples all surrogate datasets from a reference model with independent variables. The latter has been criticized based on the argument that retaining at least one factor implies that the analyzed variables share more variance than implied by the surrogate datasets in PA (Braeken & van Assen, 2017; Green et al., 2012; Ruscio & Roche, 2012; Saccenti & Timmerman, 2017; Turner, 1998). NEST is based on work by Green et al. (2012) who have proposed a 'Revised Parallel Analysis' (RPA) with which sequential conditioning of simulated sampling distributions on every retained

factor was introduced. The differences in design between NEST and RPA are mainly computational and concern the computation of the *k*-factor models to simulate surrogate datasets (for a detailed contrast between NEST and RPA, see Achim, 2017).

Previous simulation studies showed that NEST is as accurate as, or more accurate than, several PA variants for simulated datasets under a wide range of factor models (Achim, 2017; Brandenburg & Papenberg, 2022). However, in these studies NEST was tested only with continuous variables sampled from multivariate normal distributions. As will be derived in the following sections, the performance of NEST for categorical variables instead of continuous variables requires dedicated research in light of ambiguity concerning the computation of sample correlation matrices for categorical datasets.

Problems with Categorical Variables

A widely regarded problem with categorical variables is that sample product-moment correlations—hereafter referred to as Pearson correlations—underestimate the true correlations among categorical variables if these categorical variables are mere ordinal representations of continuous constructs (Garrido et al., 2013; Green et al., 2016; Lubbe, 2019).

As an alternative to Pearson correlations, sample polychoric correlations do not compute the correlation among observed categorical variables but the correlation among latent continuous variables underlying the categorical variables (Flora & Curran, 2004; Jin & Yang-Wallentin, 2017; Muthén, 1978; Olsson, 1979a). Assuming that the latent variables are normally distributed, it can be shown that maximum-likelihood estimates of polychoric correlations are asymptotically unbiased, which is an advantage over biased Pearson correlations (Lubbe, 2019). However, even if distributional assumptions are met, computing polychoric correlations instead of Pearson correlations comes at the cost of an increased standard error (i.e., instable sample correlations in repeated sampling from the population; Garrido et al., 2013; Garrido et al., 2016; Roznowski et al., 1991; Tran & Formann, 2009;

Weng & Cheng, 2017). Concerning factor analysis for categorical variables, it has been debated whether there are substantial benefits to the estimation of factor models from polychoric correlation matrices instead of Pearson correlation matrices (Flora & Curran, 2004; Garrido et al., 2016; Goretzko et al., 2021). In the present work, the focus is not on the general use of polychoric correlations in factor analysis but on the use of polychoric correlations specifically to determine the number of factors through eigenvalues of sample correlation matrices.

The original implementation of NEST, published by Achim (2017), had been developed to test NEST for continuous variables using Pearson correlations. In the original implementation surrogate datasets are simulated from a multivariate normal distribution implied by the tested *k*-factor model for Pearson correlations of the observed data. Pearson correlations are also computed for the simulated surrogate datasets. Central research questions in the present work concern how accurately NEST determines the optimal number of factors for categorical datasets when Pearson correlations are computed and whether accuracy improves when polychoric correlations are computed instead. The well-known bias of Pearson correlations on the one hand and the inflated standard error of polychoric correlations on the other suggest that one should be careful before adopting a preference for one approach over the other. Similar work in which PA with Pearson correlations was contrasted to PA with polychoric correlations has yielded mixed results (Garrido et al., 2013; Lubbe, 2019; Timmerman & Lorenzo-Seva, 2011; Cho et al., 2009; Weng & Cheng, 2005; Tran & Formann, 2009).

The differences in eigenvalues of Pearson correlation matrices and polychoric correlation matrices have implications for NEST. Simulations by Lubbe (2019) have shown these differences remarkably clearly. First, biased Pearson correlations for categorical variables result in lower signal eigenvalues (i.e., eigenvalues corresponding to factors that should be retained) compared to unbiased Pearson correlations for continuous variables. This

is relevant to NEST because a factor corresponding to a large signal eigenvalue is more likely to be retained than a factor corresponding to a small signal eigenvalue. Second, the increased standard error of polychoric correlations increases the variance of signal eigenvalues and noise eigenvalues (i.e., eigenvalues not corresponding to factors) compared to Pearson correlations for continuous variables. This, in turn, is relevant to NEST because greater dispersion of signal eigenvalues implies that the smallest signal eigenvalue is lower and therefore less likely to be retained by NEST than with Pearson correlations for continuous variables. What is more, Lubbe has shown that both the bias of Pearson correlations and the standard errors of polychoric correlations increase for categorical variables when category probabilities—which correspond to item difficulties—are asymmetric and increase even further when category probabilities vary among variables in a dataset. In contrast, the standard error of Pearson correlations and the unbiasedness of polychoric correlations were mostly insensitive to manipulated category probabilities.

From these observations it can be anticipated that NEST is more likely to underestimate the number of factors for categorical variables due to decremented statistical power compared to applications of NEST with Pearson correlations for continuous datasets. Decremented power for categorical variables can be anticipated with Pearson correlations due to bias and with polychoric correlations due to standard error. The investigations reported here also took varying category probabilities in categorical datasets into account. This was done due to the effects on eigenvalues reported by Lubbe (2019) which indicate that the performance of NEST for categorical datasets may depend not only on the computation of correlations but also on the interaction between the computation of correlations and category probabilities.

Additionally, in instances where no signal eigenvalue is rejected by NEST, varying category probabilities in a dataset also may promote overestimation of the number of factors. This potential problem can be anticipated to affect NEST when Pearson correlations are computed for categorical variables and continuous surrogate datasets as it is done in the

original implementation of NEST (Achim, 2017). Aligning category probabilities among categorical variables can act as an additional source of correlation that is not implied by a factor model but is reflected in sample Pearson correlations (Garrido et al., 2013; Green et al., 2016; Lim & Jahng, 2019; Tran & Formann, 2009; Yang & Xia, 2015). Sources of correlation related merely to category probabilities (i.e., item difficulties) are commonly referred to as *difficulty factors*. Crucially, the presence of difficulty factors can inflate noise eigenvalues from sample correlation matrices (Garrido et al., 2013; Lim & Jahng, 2019). Given that continuous surrogate datasets in the original implementation of NEST do not account for the presence of difficulty factors in the analyzed dataset, tested eigenvalues corresponding to difficulty factors may exceed their simulated sampling distribution. This would likely cause NEST to reject the null hypothesis of k-factor sufficiency erroneously due to ill-behaved simulations of sampling distributions for tested eigenvalues.

Simulation of Categorical Variables

To investigate whether Pearson correlations or polychoric correlations result in more accurate performance by NEST for categorical variables, a simulation method for categorical variables under factor models designed by Yang and Xia (2015) was adopted. Specifically, their simulation method was adopted as it includes instructions to manipulate category probabilities. The three levels of category probabilities in the design by Yang and Xi can be described as symmetric, invariant asymmetric, and varying asymmetric. These levels distinctly impact the bias of Pearson correlations and the standard errors of polychoric correlations according to Lubbe (2019) and account for difficulty factors through varying asymmetric category probabilities.

The method by Yang and Xia (2015) starts by sampling a dataset from a multivariate normal distribution with a covariance matrix that is determined by a factor model. At the population-level, the marginal distribution of each variable is the standard normal distribution. Each normal value is then transformed into a categorical value depending on

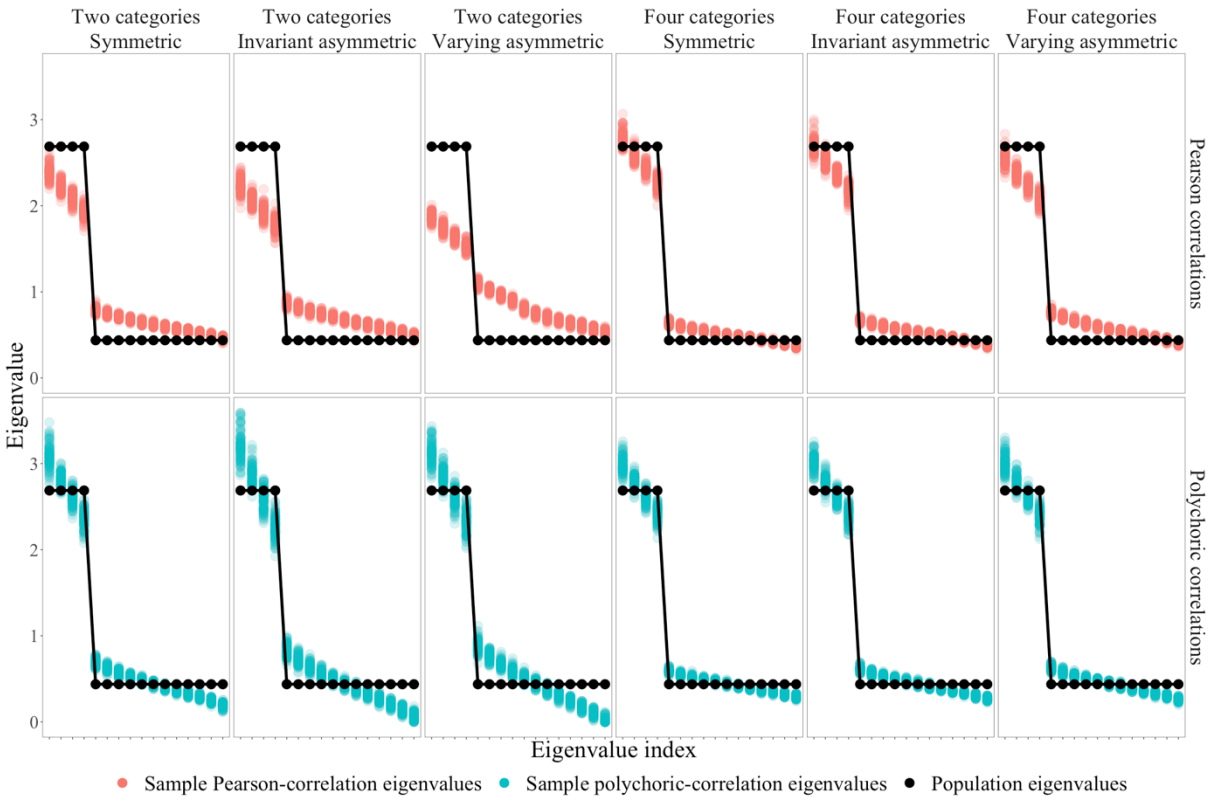
whether the normal value exceeds a predetermined threshold corresponding to intervals in the cumulative distribution function of the standard normal distribution, yielding specific category probabilities. Using threshold values in accordance to the normal distribution is common practice in simulations of categorical variables (Cho et al., 2009; Garrido et al., 2013; Green et al., 2016). Yang and Xia reported separate thresholds that transform normal variables into categorical variables with two response categories (i.e., binary variables) and variables with four response categories. The thresholds for symmetric category probabilities were {0.00} for two categories and {-1.00, 0.00, 1.00} for four categories. For invariant asymmetric category probabilities, the thresholds were {1.00} for two categories and {0.00, 0.75, 1.50} for four categories. The invariant asymmetric thresholds were used for all variables in a dataset, resulting in equally skewed category probability distributions in all variables. For varying asymmetric category probabilities, the same thresholds from the invariant asymmetric condition were used, albeit in sign-reversed form for every second variable. Following this method, symmetric category probabilities and invariant asymmetric category probabilities imply homogeneous item difficulties while varying asymmetric category probabilities imply heterogeneous item difficulties.

To illustrate how the method by Yang and Xia (2015) affects eigenvalues of Pearson correlation matrices and polychoric correlation matrices respectively, datasets with two response categories and datasets with four response categories were simulated under a factor model and compared the eigenvalues of the sample correlation matrices to the eigenvalues of the model-implied correlation matrix as population-level reference. The technical details of this simulation—and of the simulations reported in the following sections—are explained in the appendix. Figure 1 displays the eigenvalues from this simulation. With Pearson correlations, underestimations of population-level correlations in samples caused signal eigenvalues to be lower than their population-level reference. Conversely, noise eigenvalues of Pearson correlation matrices exceeded population-level noise eigenvalues. For polychoric

correlations[1], the eigenvalues in Figure 1 reflect the increased standard error in two aspects. First, the dispersion of signal eigenvalues within each index was greater compared to Pearson correlations. Second, the dispersion of all signal eigenvalues and all noise eigenvalues around their population-level reference was greater than with Pearson correlations. Figure 1 also shows that the effects of bias and standard error on sample eigenvalues were stronger (a) for asymmetric category probabilities (both invariant and varying) than for symmetric probabilities and (b) for binary variables than for variables with four response categories.

Figure 1

Sample Eigenvalues as a Function of Correlation Measure, Response Categories, And Category Probabilities



Note. The datasets in this figure were simulated under a factor model with four orthogonal factors, indicated by four variables each with loadings of .75 and without cross-loadings. Population eigenvalues are the eigenvalues of the model-implied correlation matrix. The sample size was set to $N = 500$ in all simulated datasets. The arrangement of cells constitutes between-subject manipulations. All cells summarize an independent set of datasets,

[1] Polychoric correlations for binary variables may also be referred to as tetrachoric correlations. Here, the term polychoric correlations is used for consistency with the naming for variables with four response categories.

accounting for the respective correlation measure, number of response categories, and category probabilities. A total of 100 datasets were sampled per cell.

Of course, Figure 1 has mainly illustrative purposes and does not allow firm conclusions about whether Pearson correlations or polychoric correlations are superior in NEST for categorical variables. Such firm conclusions may be derived from large-scale simulations in which different implementations of NEST with both types of correlations are applied to simulated datasets under an array of conditions.

NEST with Polychoric Correlations

In this section, a candidate implementation of NEST that builds on the estimation of polychoric correlations is proposed. Implementing a NEST variant with polychoric correlations is not trivial due to the requirement to simulate surrogate datasets. It is necessary to simulate a sampling distribution of eigenvalues under the null hypothesis of *k*-factor sufficiency. To this end, surrogate datasets need to accommodate the polychoric correlations from the analyzed dataset. Figure 1 shows that eigenvalues from datasets under factor models are sensitive to the standard errors of sample correlations. Hence, computing polychoric correlations in a dataset, then computing a *k*-factor reference model, and then adhering the original implementation by simulating continuous surrogate datasets and estimating Pearson correlations for the surrogate datasets would simulate sampling distributions of tested eigenvalues that are less dispersed than the actual sampling distributions of eigenvalues of polychoric correlation matrices. Consequently, using polychoric correlations for a dataset and Pearson correlations for continuous surrogate datasets in NEST would frequently suggest rejecting the null hypothesis when the null hypothesis is actually true. This is so because the largest noise eigenvalues of polychoric correlation matrices can be expected to exceed the

largest noise eigenvalues of Pearson correlation matrices due to the increased standard error of polychoric correlations (see Figure 1)[2].

To simulate adequate sampling distributions of eigenvalues of polychoric correlation matrices, the proposed implementation of NEST built on polychoric correlations also computes polychoric correlations for surrogate datasets. To this end, continuous variables in surrogate datasets are transformed into categorical variables in similar fashion as in the simulation of categorical variables by Yang and Xia (2015; see above). To achieve compatible standard errors of polychoric correlations between the analyzed dataset and its surrogate datasets, categorical surrogate datasets need to be simulated with category probabilities that match the probabilities of the corresponding variables from the analyzed dataset (Lubbe, 2019). In the proposed implementation of NEST, category probabilities are computed in all variables separately. Then, in a first step, continuous surrogate datasets are simulated under the multivariate normal distribution implied by the *k*-factor model. Next, for each variable in each surrogate dataset, the quantiles of the variables' simulated values are computed according to the category probabilities observed in the corresponding variable. Quantiles are used as thresholds to transform continuous surrogate datasets into categorical datasets. This approach has been suggested by A. Achim—the author of the original implementation of NEST (personal communication, June 28, 2021). Specifying thresholds for each variable in each surrogate dataset separately using its individual quantiles guarantees that all variables in surrogate datasets include the same number of response categories as the variables from which the surrogate datasets were derived. Otherwise, computation of polychoric correlations may fail in surrogate datasets due to inconsistent numbers of categories. The appendix includes technical details of the handling of the computational cost

[2] An implementation of NEST that used polychoric correlations for analyzed datasets and Pearson correlations for continuous surrogate datasets was tested in an unpublished simulation. The results confirmed that this implementation vastly overestimates the number of factors, disqualifying it from further consideration.

of computing polychoric correlations for each surrogate dataset and the handling of nondefinite correlation matrices.

Simulation Study

While the proposed implementation of NEST with polychoric correlations can be expected to simulate adequate sampling distributions of tested eigenvalues under the null hypothesis, the potential problem of decremented power compared to applications to continuous datasets due to standard error remained. Without thorough investigation in simulations, it cannot be predicted whether the potential decrement in power due to polychoric correlations deteriorates NEST's performance more severely than the potential decrement in power and sensitivity to difficulty factors due to Pearson correlations. A simulation study was conducted to compare two variants of NEST, one computing Pearson correlations for a dataset and continuous surrogate datasets and one computing polychoric correlations for a dataset and categorical surrogate datasets. The goal was to test (a) whether the anticipated problems could indeed be observed for categorical datasets and (b) how severely the performance of both variants of NEST would deteriorate. Furthermore, the two variants were compared directly to test whether polychoric correlations offer benefits to the performance of NEST for categorical variables.

Simulated Data Structures

In the present simulation study, seven independent variables were manipulated in a fully-crossed design: the true number of factors (2, 4), the number of variables per factor (4, 7), the distribution of loading parameters ($U(.40, .50)$, $U(.70, .80)$), the inter-factor correlation parameters (.20, .70), the sample size (N) of datasets (100, 500), the number of response categories (two categories, four categories), and category probabilities in variables (symmetric, invariant asymmetric, varying asymmetric). Combined, the design implied 192 conditions. For each condition 100 datasets were simulated, resulting in 19.200 simulated datasets in total.

Each condition implied a family of factor models according to Equation 1. The number of variables was the product of the number of factors and the number of variables per factor. Each factor was indicated through nonzero loadings by the number of variables per factor. Each nonzero loading parameter was independently sampled from the according uniform distribution to simulate heterogeneity of loadings on population level. Consequently, each variable only had one nonzero loading parameter, implying perfect simple-structure models (Revelle & Rocklin, 1979). All off-diagonal elements of the inter-factor correlation matrix were set to the inter-factor correlation parameter according to the simulation's design (see above). Together, these manipulations implied the term $\Lambda\Psi\Lambda^{T}$ from Equation 1, which was transformed into the model-implied correlation matrix by incrementing its main-diagonal elements to 1. It follows that only the factor models determined the population correlation matrix. There was no source of correlation at the population level other than that implied by $\Lambda\Psi\Lambda^{T}$. Given that the manipulation of the loading parameters involved random number sampling, 100 factor models per condition were generated and one dataset per factor model was simulated to achieve 100 datasets per condition (parallel to Brandenburg & Papenberg, 2022). Response categories and category probabilities were manipulated as suggested by Yang and Xia (2015). The levels of response categories and category probabilities in the simulation study were the same as those illustrated in Figure 1. All simulated variables were originally sampled from a multivariate normal distribution implied by the respective factor model and were transformed into categorical datasets according to the method by Yang and Xia.

Due to the computational costs of polychoric correlations, the present simulation study was designed with a smaller range of conditions than previous simulations that had applied NEST to continuous variables (Brandenburg & Papenberg, 2022). The levels of the independent variables that were unrelated to the categorization method by Yang and Xia (2015) were specified with the intention to avoid bottom and ceiling effects in the

performance of NEST variants. Specifically, the statistical power of NEST diminishes (a) as the number of factors and the inter-factor correlations increase, and (b) as the number of variables per factor, loading parameters, and sample size decrease (Brandenburg & Papenberg, 2022). Hence, two levels were selected for each of these independent variables to include an 'easy' and a 'difficult' level and combined them in a fully-crossed design.

Investigated Methods

The two competing NEST variants were applied to all 19,200 simulated datasets to investigate how accurately they recovered the number of factors of the factor models under which datasets had been simulated. In the following sections, the variant that used Pearson correlations will be referred to as $NEST_{Pearson}$ and the variant that used polychoric correlations will be referred to as $NEST_{poly}$. For both variants the null hypothesis of *k*-factor sufficiency was tested with 200 surrogate datasets for each test and $\alpha = .05$.

Additionally, a variant of PA was applied to all simulated datasets to provide benchmarks for the performance of NEST. PA is frequently used to add benchmarks in comparative simulations (Achim, 2017; Braeken & van Assen, 2017; Golino et al., 2020; Goretzko & Bühner, 2020; Lorenzo-Seva et al., 2011; Ruscio & Roche, 2012). Here the implementation of PA published by Lubbe (2019) was adopted. Lubbe's implementation is tailored specifically to categorical datasets: Polychoric correlations are computed for the analyzed dataset and surrogate datasets while all variables in surrogate datasets reproduce the observed category probabilities (as in the proposed $NEST_{poly}$ implementation). Lubbe conducted a simulation study and concluded that polychoric correlations and reproduced category probabilities in surrogate datasets in PA are key to optimal performance for categorical datasets. Therefore, their implementation was used in the present simulation to explore how it compares to $NEST_{Pearson}$ and $NEST_{poly}$. To highlight that this implementation of PA computes polychoric correlations, we refer to it as PA_{poly}. The software solutions to compute polychoric correlations were the same in $NEST_{poly}$ and PA_{poly} (see appendix). We set

the number of surrogate datasets in PA_{poly} to 200 for consistency with the NEST variants. As suggested by Lubbe, the threshold of reference eigenvalues which the tested eigenvalues had to exceed in order to retain the corresponding factor was the 50th percentile.

Analysis

The number of factors suggested by $NEST_{Pearson}$, $NEST_{poly}$, and PA_{poly} was recorded for all simulated datasets. As in a previous simulation study on NEST (Brandenburg & Papenberg, 2022), each solution was labeled according to one of four (exhaustive) outcomes: a solution was 'accurate' if the number of recovered factors was equal to the ground-truth number of factors form the factor model under which the analyzed dataset had been simulated. We defined the accuracy of a method as the proportion of its accurate solutions in all four possible outcomes. A solution was labeled 'overestimated' if the number of recovered factors exceeded the ground-truth number of factors and 'underestimated' if the number of recovered factors was lower than the ground-truth number of factors. An overestimated solution can be regarded a Type-1 error (i.e., the null hypothesis is rejected while testing a noise eigenvalue) and an underestimated solution can be regarded a Type-2 error (i.e., the null hypothesis was not rejected while testing a signal eigenvalue). Finally, a solution was 'undefined' if the implementation failed to converge on a solution.

Pilot Simulation

The power of $NEST_{Pearson}$ and $NEST_{poly}$ was anticipated to be decremented for categorical datasets compared to $NEST_{Pearson}$ for continuous datasets. Also, the potential presence of difficulty factors in Pearson correlation matrices for categorical datasets was anticipated to increase the risk of overestimation by $NEST_{Pearson}$ compared to applications of $NEST_{Pearson}$ for continuous datasets. Therefore, a pilot simulation was conducted in which $NEST_{Pearson}$ was applied to continuous datasets in order to obtain a baseline performance of NEST. The pilot simulation included the same manipulations of the number of factors, the

number of variables per factor, loading parameters, inter-factor correlation parameters, and sample size as the design introduced above. Consistent with main simulation, 100 datasets per condition were sampled from the model-implied multivariate normal distributions. The marginal probability distribution of all variables was approximately symmetric.

Availability

The source code of all reported simulations can be retrieved from the Open Science Repository (see https://osf.io/wb2ys/) that accompanies the current manuscript. This repository also contains the raw data from the present simulation study, scripts to replicate the analyses of the raw data, and the implementations of $NEST_{Pearson}$, $NEST_{poly}$, and PA_{poly}. Furthermore, the repository includes instructions on how to replicate the present simulation, either by re-simulating the same datasets that had been simulated in the present work or by simulating new datasets under the same conditions. The implementation of the present simulation can be adjusted to account for different sets of conditions. Also, instructions are provided to run simulations with different methods to determine the number of factors than those discussed here.

Results

The simulation indicated that the performance of $NEST_{Pearson}$, $NEST_{poly}$, and PA_{poly} were sensitive to the number of response categories, the level of the category probabilities, and sample size. These effects are particularly interesting for empirical research as the conditions were unrelated to factor models. Hence, in practice, these conditions can be assessed prior to applications of NEST. Performance is reported separately for the numbers of response categories, the level the category probabilities, and sample sizes.

Two Response Categories. Table 1 lists the proportions of outcomes for binary datasets depending on sample size and category probabilities. $NEST_{poly}$ was the most accurate method for binary datasets with $N = 500$ averaged across all category probabilities (67.2 %

accurate, 30.5 % underestimation, 2.3 % overestimation). With $N = 100$, PA_{poly} was the most accurate method averaged across all category probabilities (36.1 % accurate, 53.6 % underestimation, 10.3 % overestimation) while $NEST_{poly}$ was the least accurate method (28.3 % accurate, 70.8 % underestimation, 0.7 % overestimation, 0.3 % undefined).

Table 1

Outcomes for Binary Datasets as a Function of Sample Size and Category Probabilities

Method	Outcome	$N = 100$			$N = 500$		
		Symmetric	Invariant asymmetric	Varying asymmetric	Symmetric	Invariant asymmetric	Varying asymmetric
$NEST_{Pearson}$	Overestimation	2.8	16.1	11.5	3.7	17.1	44.1
	Accurate	42.6	32.5	18.4	75.9	58.9	21.1
	Underestimation	54.6	51.4	70.1	20.4	24.0	34.8
	Undefined	0	0	0	0	0	0
$NEST_{poly}$	Overestimation	1.0	0.5	0.6	2.2	1.6	3.0
	Accurate	39.7	26.4	18.8	77.1	64.7	59.9
	Underestimation	59.3	72.3	80.7	20.6	33.7	37.1
	Undefined	0	0.8	0	0	0	0
PA_{poly}	Overestimation	9.4	12.1	9.4	2.2	7.7	8.9
	Accurate	41.9	36.4	30.0	63.5	52.1	49.7
	Underestimation	48.6	51.5	60.6	34.3	40.2	41.4
	Undefined	0	0	0	0	0	0

Note. Percentage of outcomes in all simulated binary datasets.

In general, all methods performed best with $N = 500$ and symmetric category probabilities. As for the sample size, all methods underestimated the number of factors less frequently with $N = 500$ than with $N = 100$. The decreased Type-2 error rate with increased sample size illustrates how the power of NEST increases with sample size. With respect to category probabilities, all methods were most accurate with symmetric category probabilities, less accurate with invariant asymmetric category probabilities, and least accurate with varying asymmetric category probabilities. Table 1 indicates that reduced accuracy with the asymmetric category probability levels in $NEST_{poly}$ can be attributed to underestimations and not to overestimations. $NEST_{poly}$ rarely overestimated the number of factors and did not exceed its normative Type-1 error rate of 5 % (implied by its α level). In contrast, asymmetric

category probabilities simultaneously caused more underestimations and overestimations by $NEST_{Pearson}$. The only exception was that $NEST_{Pearson}$ showed less underestimations with invariant asymmetric category probabilities than with symmetric probabilities with $N = 100$. Overall, this indicates that $NEST_{Pearson}$ likely suffered from reduced power as well as sensitivity to difficulty factors. A striking problem with $NEST_{Pearson}$ was that it overestimated the number of factors more frequently with $N = 500$ than with $N = 100$. Table 1 shows that overestimations by $NEST_{Pearson}$ were particularly frequent with $N = 500$ and varying asymmetric category probabilities.

In comparison, $NEST_{poly}$ underestimated the number of factors more frequently than $NEST_{Pearson}$ with all sample sizes and category probabilities. Hence, the power of $NEST_{poly}$ was considerably lower than the power of $NEST_{Pearson}$ for binary datasets in the present simulation.

Four Response Categories. Table 2 lists the proportions of outcomes for datasets with four response categories. $NEST_{Pearson}$ was the most accurate method on average across all category probabilities with $N = 500$ (70.7 % accurate, 15.0 % underestimation, 14.2 % overestimation) and with $N = 100$ (53.1 % accurate, 43.7 % underestimation, 3.1 % overestimation).

Table 2

Outcomes for Datasets with Four Response Categories as a Function of Sample Size and Category Probabilities

Method	Outcome	$N = 100$			$N = 500$		
		Symmetric	Invariant asymmetric	Varying asymmetric	Symmetric	Invariant asymmetric	Varying asymmetric
$NEST_{Pearson}$	Overestimation	1.3	3.1	5.1	1.7	4.1	37.0
	Accurate	59.1	53.7	46.7	85.1	80.7	46.2
	Underestimation	39.7	43.2	48.2	13.2	15.2	16.8
	Undefined	0	0	0	0	0	0
$NEST_{poly}$	Overestimation	0	0	0.1	0	0	0
	Accurate	25.6	30.0	29.4	45.1	70.4	45.2
	Underestimation	74.4	70.0	70.5	54.9	29.6	54.8
	Undefined	0	0	0	0	0	0
PA_{poly}	Overestimation	5.8	7.0	8.3	0	0.4	0.2
	Accurate	48.9	46.3	46.2	70.4	68.1	68.6
	Underestimation	45.3	46.7	45.5	29.6	31.5	31.2
	Undefined	0	0	0	0	0	0

Note. Percentage of outcomes in all simulated datasets with four response categories.

Compared to binary datasets, $NEST_{Pearson}$ and PA_{poly} improved ($NEST_{Pearson}$: 41.6 % accurate for all binary datasets, 61.9 % accurate for all datasets with four categories; PA_{poly}: 45.6 % accurate for all binary datasets, 58.1 % accurate for all datasets with four categories) while $NEST_{poly}$ became less accurate (47.8 % accurate for all binary datasets, 41.0 % accurate for all datasets with four categories). In fact, $NEST_{poly}$ was the least accurate method for four categories by far.

The effects of sample size and category probabilities evident in Table 2 are similar to the effects of sample size and category probabilities from binary datasets in that methods mostly benefitted from increased sample size and symmetric category probabilities. An exception is that $NEST_{poly}$ achieved its highest accuracy with $N = 500$ and invariant asymmetric category probabilities and was most likely to underestimate the number of factors with symmetric category probabilities. The low proportions of overestimations and the high proportions of underestimations by $NEST_{poly}$ for four categories suggest that the inaccuracy of $NEST_{poly}$ can be attributed to a severe lack of power. The inflation of the Type-1 error rate in

$NEST_{Pearson}$ substantially exceeded its normative Type-1 error rate of 5 % only with $N = 500$ and varying asymmetric category probabilities. Crucially, as with binary datasets, results indicate that $NEST_{Pearson}$ was more prone to overestimation with $N = 500$ than with $N = 100$—particularly with varying asymmetric category probabilities. No NEST variant outperformed PA_{poly} for datasets with four categories with varying asymmetric category probabilities.

Pilot Simulation. All results for categorical datasets can be put into perspective by comparing them to the performance of $NEST_{Pearson}$ for continuous datasets in the pilot simulation. For continuous datasets, $NEST_{Pearson}$ performed better with $N = 500$ (88.8 % accurate, 10.0 % underestimation, 1.2 % overestimation) than with $N = 100$ (63.1 % accurate, 35.0 % underestimation, 1.9 % overestimation).

The proportions of outcomes in Table 1 indicate that all methods performed worse for binary datasets than $NEST_{Pearson}$ for continuous datasets. For binary datasets, $NEST_{Pearson}$ and $NEST_{poly}$ underestimated the number of factors more frequently throughout all category probabilities with and $N = 500$ and $N = 100$ respectively. This indicates $NEST_{Pearson}$ and $NEST_{poly}$ indeed suffered from decremented power compared to $NEST_{Pearson}$ for continuous datasets, albeit $NEST_{poly}$ more so than $NEST_{Pearson}$. However, overestimations by $NEST_{Pearson}$ with asymmetric category probabilities further contributed to its decrement in accuracy for binary datasets compared to continuous datasets.

While $NEST_{Pearson}$ improved for datasets with four categories, its performance was still worse than for continuous datasets. Compared to its performance for continuous datasets with $N = 500$ and $N = 100$ respectively, $NEST_{Pearson}$ showed more underestimations with all category probabilities. With varying asymmetric category probabilities, overestimations by $NEST_{Pearson}$ were also more frequent than for continuous variables. This pattern is similar to the results for binary datasets. It follows that reduced accuracy of $NEST_{Pearson}$ for categorical datasets compared to continuous datasets can be attributed to more frequent underestimation

with all category probabilities and—simultaneously—more frequent overestimation with varying asymmetric category probabilities.

Discussion

A key motivation to investigate $NEST_{Pearson}$ and $NEST_{poly}$ in a simulation was to test whether the anticipated problems for categorical datasets would deteriorate their performance compared to application to continuous datasets. The results from the present simulated confirm that $NEST_{Pearson}$ and $NEST_{poly}$ performed worse for categorical datasets than $NEST_{Pearson}$ for continuous datasets.

Statistical Power of NEST

There was reason to anticipate that the underestimation of model-implied correlations through sample Pearson correlations and the large standard error of sample polychoric correlations would result in deflated signal eigenvalues (see Figure 1). The frequent underestimation of the optimal number of factors by $NEST_{Pearson}$ and $NEST_{poly}$ in the present simulation study implies that signal eigenvalues were indeed often too small to exceed their simulated sampling distribution when the null hypothesis of *k*-factor sufficiency was incorrect. Which NEST variant suffers from a stronger decrement in power, was an open question which could not be answered without simulations. Interestingly, the stronger tendency toward overestimation by $NEST_{poly}$ compared to $NEST_{Pearson}$ for categorical datasets suggests that polychoric correlations reduced the statistical power of NEST more than Pearson correlations despite the unbiasedness of polychoric correlations.

Furthermore, it was anticipated that the power of $NEST_{Pearson}$ and $NEST_{poly}$ depends on category probabilities. $NEST_{Pearson}$ and $NEST_{poly}$ mostly underestimated the number of factors more frequently with invariant or varying asymmetric category probabilities than with symmetric category probabilities. This is in line with reports that the bias of Pearson correlations and the standard error of polychoric correlations are larger with asymmetric category probabilities than with symmetric probabilities (Lubbe, 2019), which is also

reflected in eigenvalues of the respective sample correlation matrices (see Figure 1). Inconsistent with this pattern, $NEST_{Pearson}$ underestimated the number of factors more often for binary datasets with $N = 100$ and symmetric category probabilities than with invariant asymmetric category probabilities. Also inconsistent with this pattern, $NEST_{poly}$ underestimated the number of factors for datasets with four categories most frequently with symmetric category probabilities. Figure 1 provides no explanations for these inconsistencies, which indicates that the dependence of NEST variants on the number of response categories and category probabilities was more complex than anticipated and requires further research. Still, the present results add support to the notion that asymmetric category probabilities in categorical datasets obstructs factor retention.

Difficulty Factors

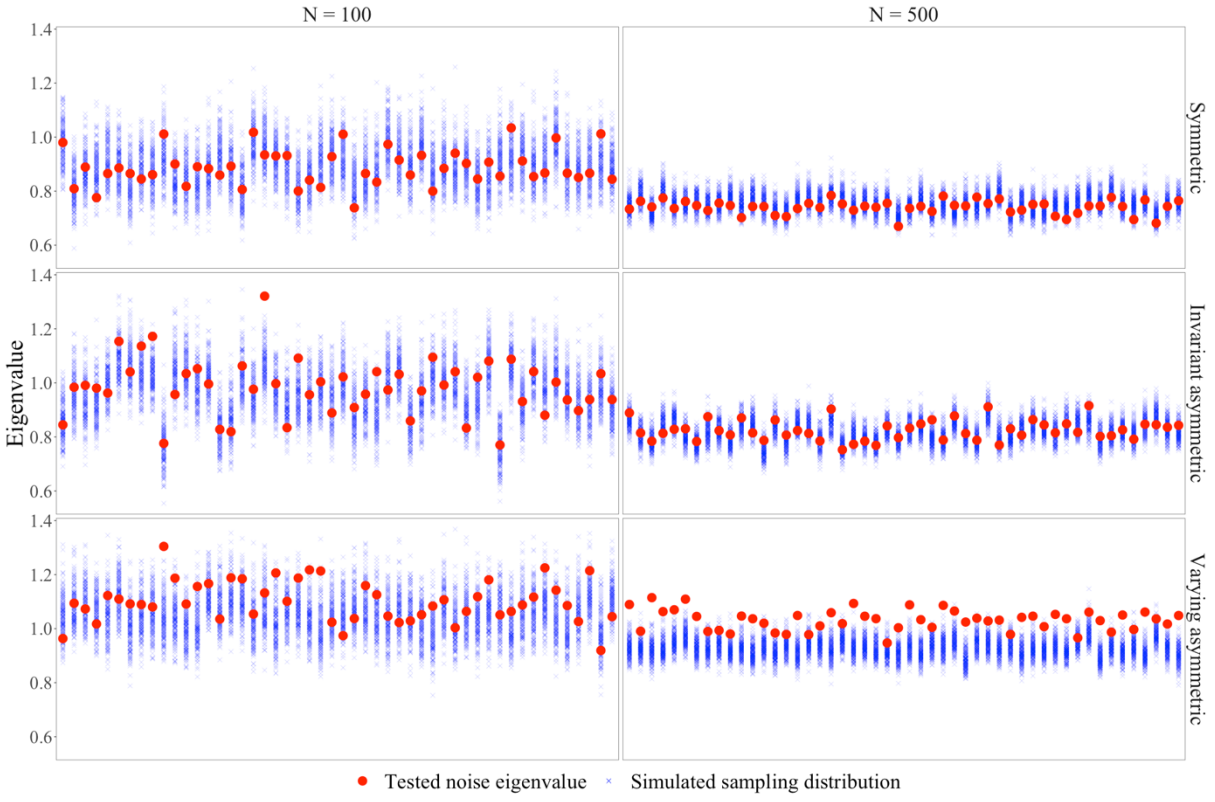
Another problem with $NEST_{Pearson}$ related to category probabilities was its alarming tendency to overestimate the number of factors for datasets with varying asymmetric category probabilities. This result fits the premise that difficulty factors occur in datasets when item pairs of similar difficulty show substantial Pearson correlations (Garrido et al., 2013; Green et al., 2016; Lim & Jahng, 2019; Tran & Formann, 2009; Yang & Xia, 2015) and that the continuous datasets simulated in $NEST_{Pearson}$ failed to account for category probabilities as a confounded source of correlation. Thus, the tendency toward overestimation by $NEST_{Pearson}$ provides evidence that the original implementation—which was not designed for application to categorical variables—indeed fails to safeguard against retention of difficulty factors.

Overestimations by $NEST_{Pearson}$ occurred more frequently with $N = 500$ than with $N = 100$. For a method to infer the number of factors, worse performance with supposedly clearer indication of population-level properties is highly undesirable (cf. Garrido et al., 2013). Two explanations for the inflated Type-1 error rate with increasing sample size come to mind: First, $NEST_{Pearson}$ showed fewer underestimations with $N = 500$ than with $N = 100$, thus exposing $NEST_{Pearson}$ to noise eigenvalues more often. Second, with $N = 500$, the simulated

sampling distribution of the eigenvalues in $NEST_{Pearson}$ may have deteriorated further than with $N = 100$. To better understand the causes of overestimation, additional datasets were simulated under the conditions of the simulation presented above. All eigenvalues tested and simulated by $NEST_{Pearson}$ were stored. Figure 2 contrasts the largest noise eigenvalues from datasets with $N = 500$ and $N = 100$ to their simulated sampling distributions from continuous surrogate datasets with matching sample size. Eigenvalues in $NEST_{Pearson}$ strikingly exceeded their simulated sampling distribution for datasets with $N = 500$ and varying asymmetric category probabilities. This aligns with the present simulation in which $NEST_{Pearson}$ overestimated the number of factors particularly often with $N = 500$ and varying asymmetric category probabilities. In conclusion, Figure 2 shows a localized flaw in $NEST_{Pearson}$, namely that the simulated sampling distribution of tested eigenvalues failed to accommodate varying asymmetric category probabilities, deteriorating further with increasing sample size.

Figure 2

Noise Eigenvalues and their Simulated Sampling Distributions as a Function of Sample Size and Category Probabilities



Note. The red dots indicate the first noise eigenvalue that were tested by $NEST_{Pearson}$ in a series of simulated binary datasets under a factor model with two orthogonal factors indicated by four variables each with loadings of .75. The blue crosses indicate the respective sampling distributions of eigenvalues under the null hypothesis. Each vertical segment within the six cells represents an application of NEST to an independently simulated dataset. In total, 50 datasets per cell were sampled. Given that the red dots indicate noise eigenvalues, the red dots should have the same distributions as the blue crosses. With varying asymmetric category probabilities and $N = 500$, the red dots clearly outrank the blue crosses, which reveals that sampling distributions simulated by $NEST_{Pearson}$ failed to accommodate the tested noise eigenvalues.

Empirical Example. Lubbe (2019) applied PA_{poly} to an empirical dataset with binary variables to test whether polychoric correlations and reproduced category probabilities in surrogate datasets prevent sensitivity to difficulty factors in PA. Here, their analysis was replicated with $NEST_{Pearson}$ and $NEST_{poly}$. The dataset is a sample ($N = 150$) of Bond's Logical Operations Test, which includes 35 binary variables (Bond & Fox, 2007, as cited by Revelle, 2022); it is available in the R package *psychTools* (Version 2.2.5; Revelle, 2022; retrievable as *psychTools::blot*) as a toy dataset in the context of item response theory. As such, the items are assumed to reflect a single common factor. The items' mean difficulty is .75 ($SD = 0.13$) with two item difficulties below .50 (i.e., .36, .49). Hence, the dataset can be considered in between the category probability levels 'invariant asymmetric' and 'varying asymmetric' of the present simulation. Consistent with the present simulation, $NEST_{poly}$ was accurate for the binary dataset and suggested one factor while $NEST_{Pearson}$ was inaccurate and suggested three factors. This example emphasizes that $NEST_{Pearson}$ tends to overestimate the number of factors for categorical datasets with asymmetric category probabilities, which should be kept in mind in empirical applications.

Recommendations

The goal of the research reported here was to compare $NEST_{Pearson}$ and $NEST_{poly}$ to explore potential guidelines about which variant should be preferred for categorical datasets. The present simulation was deliberately designed to simulate situations in which the

anticipated problems of $NEST_{Pearson}$ and $NEST_{poly}$ should be easily observed, facilitating an assessment of their relative differences. More importantly, however, the low accuracies obvious from Tables 1 and 2 indicate that neither $NEST_{Pearson}$ nor $NEST_{poly}$ achieved satisfactory performance under several conditions in the present simulation.

With $N = 100$, no method reached 50 % accuracy for binary datasets and no method reached 60 % accuracy for datasets with four response categories. Based on the present results, $N = 500$ is recommended as the minimum sample size to tackle the number-of-factors problem in categorical datasets with four or fewer response categories. The levels of sample size (100; 500) in the present simulation do not justify recommendation of a lower minimum sample size.

With $N = 500$, in relative terms, $NEST_{poly}$ outperformed $NEST_{Pearson}$ for binary datasets: $NEST_{poly}$ was more accurate than $NEST_{Pearson}$ and it remained within its normative Type-1 error rate while the Type-1 error rate of $NEST_{Pearson}$ was strongly inflated with asymmetric category probabilities. Hence, $NEST_{poly}$ seems to be the preferable NEST variant for binary datasets with $N = 500$. However, given that $NEST_{Pearson}$ was substantially more accurate than $NEST_{poly}$ for datasets with four categories, the benefits of polychoric correlations to NEST outweigh their costs only for binary datasets.

For categorical datasets with more than two response categories, $NEST_{Pearson}$ seems to be the preferred NEST variant. As the bias of Pearson correlations decreases with an increasing number of response categories (see Figure 1; Green et al., 2016), it can be assumed that $NEST_{Pearson}$ further improves with more than four response categories. However, since $NEST_{Pearson}$ failed to outperform PA_{poly} for datasets with four categories and varying asymmetric category probabilities, recommendation of $NEST_{Pearson}$ for categorical datasets with more than two response categories should be limited to homogeneous item difficulties in light of its sensitivity to difficulty factors. For datasets with more than two categories, more research is required to develop a method that is as robust against varying item difficulties as

PA_{poly} while also improving on theorical flaws of PA (see Braeken & van Assen, 2017; Green et al., 2012; Ruscio & Roche, 2012; Saccenti & Timmerman, 2017; Turner, 1998).

The observed errors by all methods investigated in the present simulation highlight the need for guidelines to qualify a suggested number of factors as optimal in applied research where—unlike in simulations—no ground-truth typically exists. Such guidelines are particularly important for categorical datasets given that categorical datasets promoted underestimations *and* overestimations by NEST, depending on the implementation. In general, the optimal solution to the number-of-factors problem does not miss factors of substantive importance and does not retain factors than suit no sound interpretation (Braeken & van Assen, 2017; Preacher et al., 2013).

Underestimations by NEST indicate that failure to reject the null hypothesis of k-factor sufficiency for a tested eigenvalue does not imply that the eigenvalue corresponds to a negligible factor. A guideline to avoid missing a factor with NEST is to employ large samples to increase its statistical power, which is supported by the present simulation ($N \geq 500$ for categorical datasets). When large samples are infeasible, the number of factors in EFA may be increased as long as the added factors provide contributions to the factor model that are deemed substantial according to the interpretation of model parameters in light of domain-specific theory.

What is more, overestimations by NEST indicate that statistical significance of an eigenvalue in NEST does not imply that the eigenvalue corresponds to substantive contribution by a distinct factor (Brandenburg & Papenberg, 2022). Known alternative explanations for significance in NEST are mere sampling variance (i.e., a Type-1 error), the unaccounted presence of a difficulty factor, or cumulative contributions of minor sources of correlations that do not correspond to a factor (Achim, 2021; Auerswald & Moshagen, 2019; Cosemans et al., 2022; Lim & Jahng, 2019). These alternative explanations may also lead one

to consider numbers of factors below the solution of NEST when not all factors accepted by NEST are interpretable post rotation (see Fabrigar et al., 1999).

A strategy to avoid relying on potentially erroneous solutions suggested in multiple recent publications is to consider solutions of different methods (Auerswald & Moshagen, 2019; Goretzko et al., 2021; Preacher et al., 2013). Combining different solutions requires that the relative performance of the respective methods is well-understood (Li et al., 2020) in order to interpret conflicting solutions of different methods. In the present simulation, heterogeneous item difficulties in categorical datasets caused frequent overestimations by $NEST_{Pearson}$ and frequent underestimations by $NEST_{poly}$. Following this observation, it was tested whether a combination rule of $NEST_{Pearson}$ and $NEST_{poly}$ would improve accuracy with heterogeneous item difficulties by treating the solution by $NEST_{poly}$ as the lower bound for the number of factors and the solution by $NEST_{Pearson}$ as the upper bound. To this end, only datasets from the present simulation with varying asymmetric category probabilities and converged solutions by $NEST_{Pearson}$ and $NEST_{poly}$ were considered. The boundaries covered the true number of factors in 29.4 % of binary datasets with $N = 100$, in 62.5 % of binary datasets with $N = 500$, in 51.7 % of datasets with four categories with $N = 100$, and in 83.3 % of datasets with four categories and $N = 500$. Compared to the accuracies listed in Tables 1 and 2, the boundaries covered the true number of factors more often than it had been hit by $NEST_{Pearson}$ and $NEST_{poly}$ individually. An obvious problem with this combination rule is that varying asymmetric category probabilities not only promoted overestimations by $NEST_{Pearson}$ but also underestimations. Hence, $NEST_{Pearson}$ should not be expected to provide a reliable upper bound for the optimal number of factors. Still, given that factor retention for categorical variables with heterogeneous item difficulties remains challenging for NEST and PA alike, this combination rule may be a useful heuristic that exploits shortcomings of the individual methods.

The current comparison of $NEST_{Pearson}$ and $NEST_{poly}$ also has implications for simulation studies aimed at investigating methods to determine the number of factors similar to the present study. For instance, Lim and Jahng (2019) compared traditional PA to RPA—which is highly similar to NEST—in a simulation that included datasets with four response categories. Their tested implementation of RPA computed polychoric correlations for the analyzed dataset and for categorized surrogate datasets, similar to the present implementation of $NEST_{poly}$. In light of the present simulation, which limits benefits of polychoric correlations for NEST to binary datasets, Lim and Jahng likely tested a suboptimal implementation of RPA. We suggest that further simulation studies which include NEST or RPA as well as categorical datasets should include a variant that computes Pearson correlations even for categorical datasets.

Limitations

A limitation of the present work is that the simulation was designed with a restricted set of conditions that did not target optimal conditions for NEST and PA. Given that both NEST and PA are eigenvalue-based methods, their performance can be expected to improve for categorical datasets when factors explain more variance across all variables. For Pearson correlation matrices and polychoric correlation matrices, signal eigenvalues are larger, for instance, in datasets with more variables per factor (Auerswald & Moshagen, 2019). However, it requires further simulations to test whether there are conditions that nullify the identified problems with $NEST_{Pearson}$ and $NEST_{poly}$ for categorical datasets.

The present simulation may be considered idealistic in that it did not account for cross-loadings of simulated variables on multiple factors (Brandenburg & Papenberg, 2022; Li et al., 2020). Given that the number-of-factors problem implies that it is unknown how many factors determine the variables, it should not be assumed in exploratory factor analysis that variables are determined by one common factor each (Achim, 2020). Model-implied correlation matrices of factor models with substantial cross-loadings in the majority of

variables can yield lower signal eigenvalues (excluding the largest signal eigenvalue) than correlation matrices of factor models without cross-loadings (Brandenburg & Papenberg, 2022). Therefore, future research targeting categorical datasets should assess to what extent categorical variables with cross-loadings deteriorate the performance of NEST.

Furthermore, the present work only accounts for categorical datasets in which all variables have the same number of response categories. The present results do not generalize to datasets with mixed scales. Further research is needed to address optimal estimation of correlation in NEST for mixed datasets.

Finally, another limitation concerns the assumption of normally distributed latent variables by the maximum-likelihood estimator of polychoric correlations (Olsson, 1979a). Whenever polychoric correlations were computed (i.e., Figure 1, $NEST_{poly}$, PA_{poly}), categorical variables had been simulated by transforming normally distributed variables using predetermined thresholds (Yang & Xia, 2015). Therefore, the distributional assumptions of the estimator of polychoric correlations were never violated. However, assuming that categorical variables are discrete indicators specifically of normally distributed continuous variables may not hold in empirical applications. As an example, Jin and Yang-Wallentin (2017) pointed out that income as an indicator of social-economic status may not be normally distributed due to its natural lower bound but still may be measured in ordered categories. Jin and Yang-Wallentin further showed in simulations that polychoric correlations assuming normal distribution systematically underestimate correlations between categorical variables when they are discrete transformations of continuous distributions other than the normal distribution. In our implementation of $NEST_{poly}$, the simulated categorical surrogate datasets also met distributional assumptions of polychoric correlations by design. It follows that polychoric correlations for an analyzed dataset and for surrogate datasets in $NEST_{poly}$ do not necessarily share the property of unbiasedness if the analyzed dataset unilaterally violates distributional assumptions. The present work provides no indication of the performance of

$NEST_{poly}$ when distributional assumptions of polychoric correlations are violated. Therefore, further research is required to test whether the simulated sampling distributions of tested eigenvalues in $NEST_{poly}$ remain appropriate in the presence of violated distributional assumptions.

Conclusion

All in all, the present work shows that the performance of NEST for categorical variables depends on properties of computed correlations (i.e., bias, standard error). The present simulation provides evidence that polychoric correlations for analyzed datasets and categorical surrogate datasets benefit the performance of NEST in retaining the optimal number of factors for binary datasets. However, the tested implementation of NEST using polychoric correlations required large samples to achieve satisfactory performance ($N \geq 500$) and the benefits of polychoric correlations did not extend to categorical datasets with more than two response categories per variable. For datasets with four response categories, the problems of polychoric correlations were more severe than the problems of Pearson correlations. In general, the present simulation suggests that factor retention is more error-prone for categorical datasets than for continuous datasets. More research addressing categorical variables is required to investigate which method is optimal under which condition and how potentially suboptimal solutions are handled in empirical applications of exploratory factor analysis.

Declarations

Funding: The author has no relevant financial or nonfinancial interests to disclose.

Conflicts of interest: The author has no competing interests to declare that are relevant to the content of this article.

Ethics approval: Not applicable.

Consent to participate: Not applicable.

Consent for publication: Not applicable.

Availability of data and materials: The raw results of the current simulation study are available in the Open Science Foundation repository: https://osf.io/wb2ys/

Code availability: The source code to all reported analyses, all investigated methods, and all reported simulations (including the ones to generate the displayed figures) are available in the Open Science Foundation repository: https://osf.io/wb2ys/

Acknowledgements: I thank Axel Buchner for his feedback on the current manuscript.

References

Achim, A. (2017). Testing the number of required dimensions in exploratory factor analysis. *The Quantitative Methods for Psychology*, *13*(1), 64-74. https://doi.org/10.20982/tqmp.13.1.p064

Achim, A. (2020). Esprit et enjeux de l'analyse factorielle exploratoire. [Spirit and issues of exploratory factor analysis.] *The Quantitative Methods for Psychology*, *16*(4): 213-247. https://doi.org/10.20982/tqmp.16.4.p213

Achim, A. (2021). Determining the number of factors using parallel analysis and its recent variants: Comment on Lim and Jahng (2019). *Psychological Methods*, *26*(1), 79-73. https://doi.org/10.1037/met0000269

Auerswald, M., & Moshagen, M. (2019). How to determine the number of factors to retain in exploratory factor analysis: A comparison of extraction methods under realistic conditions. *Psychological Methods*, *24*(4), 468-491. https://doi.org/10.1037/met0000200

Braeken, J., & van Assen, M. A. (2017). An empirical Kaiser criterion. *Psychological Methods*, *22*(3), 450-466. https://doi.org/10.1037/met0000074

Brandenburg, N. (2022). Factor retention in ordered categorical variables: Benefits and costs of polychoric correlations in eigenvalue-based testing [Open Science Framework Repository]. Retrieved from https://osf.io/wb2ys/

Brandenburg, N., & Papenberg, M. (2022). Reassessment of innovative methods to determine the number of factors: A simulation-based comparison of Exploratory Graph Analysis and Next Eigenvalue Sufficiency Test. *Psychological Methods*. Advance online publication. https://doi.org/10.1037/met0000527

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, *1*(2), 245-276. https://doi.org/10.1207/s15327906mbr0102_10

Cho, S. J., Li, F., & Bandalos, D. (2009). Accuracy of the parallel analysis procedure with polychoric correlations. *Educational and Psychological Measurement*, *69*(5), 748-759. https://doi.org/10.1177/0013164409332229

Conway, J. M., & Huffcutt, A. I. (2003). A review and evaluation of exploratory factor analysis practices in organizational research. *Organizational Research Methods*, *6*(2), 147-168. https://doi.org/10.1177/109442810325154

Cosemans, T., Rosseel, Y., & Gelper, S. (2022). Exploratory Graph Analysis for Factor Retention: Simulation Results for Continuous and Binary Data. *Educational and Psychological Measurement*, *82*(5), 880-910. https://doi.org/10.1177/00131644211059089

Fabrigar, L. R., Visser, P. S., & Browne, M. W. (1997). Conceptual and methodological issues in testing the circumplex structure of data in personality and social psychology. *Personality and Social Psychology Review*, *1*(3), 184-203. https://doi.org/10.1207/s15327957pspr0103_1

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, *4*(3), 272-299. https://doi.org/10.1037/1082-989X.4.3.272

Fava, J. L., & Velicer, W. F. (1992). The effects of overextraction on factor and component analysis. *Multivariate Behavioral Research*, *27*(3), 387-415. https://doi.org/10.1207/s15327906mbr2703_5

Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods, 9*(4), 466–491. https://doi.org/10.1037/1082-989X.9.4.466

Garrido, L. E., Abad, F. J., & Ponsoda, V. (2013). A new look at Horn's parallel analysis with ordinal variables. *Psychological Methods*, *18*(4), 454-474. https://doi.org/10.1037/a0030005

Garrido, L. E., Abad, F. J., & Ponsoda, V. (2016). Are fit indices really fit to estimate the number of factors with categorical variables? Some cautionary findings via Monte Carlo simulation. *Psychological Methods*, *21*(1), 93-111. https://doi.org/10.1037/met0000064

Genz, A., Bretz, F., Miwa, T., Mi, X., Leisch, F., Scheipl, F., & Hothorn, T. (2021). *mvtnorm: Multivariate normal and t distributions*. Retreived from http://cran.r-project.org/package=mvtnorm

Golino, H. F., & Epskamp, S. (2017). Exploratory graph analysis: A new approach for estimating the number of dimensions in psychological research. *PloS One*, *12*(6), e0174035. https://doi.org/10.1371/journal.pone.0174035

Golino, H., Shi, D., Christensen, A. P., Garrido, L. E., Nieto, M. D., Sadana, R., ... & Martinez-Molina, A. (2020). Investigating the performance of exploratory graph analysis and traditional techniques to identify the number of latent factors: A simulation and tutorial. *Psychological Methods, 25*(3), 292-320. https://doi.org/10.1037/met0000255

Goretzko, D., & Bühner, M. (2020). One model to rule them all? Using machine learning algorithms to determine the number of factors in exploratory factor analysis. *Psychological Methods*, *25*(6), 776–786. https://doi.org/10.1037/met0000262

Goretzko, D., Pham, T. T. H., & Bühner, M. (2021). Exploratory factor analysis: Current use, methodological developments and recommendations for good practice. *Current Psychology*, *40*(7), 3510-3521. https://doi.org/10.1007/s12144-019-00300-2

Green, S. B., Levy, R., Thompson, M. S., Lu, M., & Lo, W. J. (2012). A proposed solution to the problem with using completely random data to assess the number of factors with parallel analysis. *Educational and Psychological Measurement*, *72*(3), 357-374. https://doi.org/10.1177/0013164411422252

Green, S. B., Redell, N., Thompson, M. S., & Levy, R. (2016). Accuracy of revised and traditional parallel analyses for assessing dimensionality with binary data. *Educational and Psychological Measurement*, *76*(1), 5-21. https://doi.org/10.1177/0013164415581898

Guttman, L. (1954). Some necessary conditions for common-factor analysis. *Psychometrika*, *19*(2), 149-161. https://doi.org/10.1007/BF02289162

Henson, R. K., & Roberts, J. K. (2006). Use of exploratory factor analysis in published research: Common errors and some comment on improved practice. *Educational and Psychological Measurement*, *66*(3), 393-416. https://doi.org/10.1177/0013164405282485

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*(2), 179-185. https://doi.org/10.1007/BF02289447

Jin, S., & Yang-Wallentin, F. (2017). Asymptotic robustness study of the polychoric correlation estimation. *Psychometrika*, *82*(1), 67-85. https://doi.org/10.1007/s11336d016-9512-2

Li, Y., Wen, Z., Hau, K. T., Yuan, K. H., & Peng, Y. (2020). Effects of cross-loadings on determining the number of factors to retain. *Structural Equation Modeling: A Multidisciplinary Journal*, *27*(6), 841-863. https://doi.org/10.1080/10705511.2020.1745075

Lim, S., & Jahng, S. (2019). Determining the number of factors using parallel analysis and its recent variants. *Psychological Methods, 24*(4), 452–467. https://doi.org/10.1037/met0000230

Lorenzo-Seva, U., Timmerman, M. E., & Kiers, H. A. (2011). The Hull method for selecting the number of common factors. *Multivariate Behavioral Research*, *46*(2), 340-364. https://doi.org/10.1080/00273171.2011.564527

Lubbe, D. (2019). Parallel analysis with categorical variables: Impact of category probability proportions on dimensionality assessment accuracy. *Psychological Methods, 24*(3), 339–351. https://doi.org/10.1037/met0000171

Muthén, B. (1978). Contributions to factor analysis of dichotomous variables. *Psychometrika*, *43*(4), 551-560. https://doi.org/10.1007/BF02293813

O'Leary-Kelly, S. W., & Vokurka, R. J. (1998). The empirical assessment of construct validity. *Journal of Operations Management*, *16*(4), 387-405. https://doi.org/10.1016/S0272-6963(98)00020-5

Olsson, U. (1979). Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika*, *44*(4), 443-460. https://doi.org/10.1007/BF02296207

Preacher, K. J., Zhang, G., Kim, C., & Mels, G. (2013). Choosing the optimal number of factors in exploratory factor analysis: A model selection perspective. *Multivariate Behavioral Research*, *48*(1), 28-56. https://doi.org/10.1080/00273171.2012.710386

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retreived from https://www.r-project.org/

Revelle, W. (2022). *Psych: Procedures for psychological, psychometric, and personality research*. Retrieved from https://cran.r-project.org/package=psych

Revelle, W. (2022) *psychTools: Tools to accompany the 'psych' package for psychological research*. Retrieved from https://cran.r-project.org/package=psychTools

Revelle, W., & Rocklin, T. (1979). Very simple structure: An alternative procedure for estimating the optimal number of interpretable factors. *Multivariate Behavioral Research*, *14*(4), 403-414. https://doi.org/10.1207/s15327906mbr1404_2

Roznowski, M., Tucker, L. R., & Humphreys, L. G. (1991). Three approaches to determining the dimensionality of binary items. *Applied Psychological Measurement*, *15*(2), 109-127. https://doi.org/10.1177/014662169101500201

Ruscio, J., & Roche, B. (2012). Determining the number of factors to retain in an exploratory factor analysis using comparison data of known factorial structure. *Psychological Assessment*, *24*(2), 282-292. https://doi.org/10.1037/a0025697

Saccenti, E., & Timmerman, M. E. (2017). Considering Horn's parallel analysis from a random matrix theory point of view. *Psychometrika*, *82*(1), 186-209. https://doi.org/10.1007/s11336-016-9515-z

Schmitt, T. A. (2011). Current methodological considerations in exploratory and confirmatory factor analysis. *Journal of Psychoeducational Assessment*, *29*(4), 304-321. https://doi.org/10.1177/0734282911406653

Timmerman, M. E., & Lorenzo-Seva, U. (2011). Dimensionality assessment of ordered polytomous items with parallel analysis. *Psychological Methods*, *16*(2), 209-220. https://doi.org/10.1037/a0023353

Tran, U. S., & Formann, A. K. (2009). Performance of parallel analysis in retrieving unidimensionality in the presence of binary data. *Educational and Psychological Measurement*, *69*(1), 50-61. https://doi.org/10.1177/0013164408318761

Turner, N. E. (1998). The effect of common variance and structure pattern on random data eigenvalues: Implications for the accuracy of parallel analysis. *Educational and Psychological Measurement*, *58*(4), 541-568. https://doi.org/10.1177/0013164498058004001

Weng, L. J., & Cheng, C. P. (2005). Parallel analysis with unidimensional binary data. *Educational and Psychological Measurement*, *65*(5), 697-716. https://doi.org/10.1177/0013164404273941

Weng, L. J., & Cheng, C. P. (2017). Is categorization of random data necessary for parallel analysis on Likert-type data? *Communications in Statistics-Simulation and Computation*, *46*(7), 5367-5377. https://doi.org/10.1080/03610918.2016.1154154

Widaman, K. F. (2018). On common factor and principal component representations of data: Implications for theory and for confirmatory replications. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*(6), 829-847. https://doi.org/10.1080/10705511.2018.1478730

Yang, Y., & Xia, Y. (2015). On the number of factors to retain in exploratory factor analysis for ordered categorical data. *Behavior Research Methods*, *47*(3), 756-772. https://doi.org/10.3758/s13428-014-0499-2

Ziegler, M., & Hagemann, D. (2015*)*. Testing the unidimensionality of items. *European Journal of Psychological Assessment, 31*(4), 231-237. https://doi.org/10.1027/1015-5759/a000309

Appendix

Implementation Details

All simulations and analyses in the present work were conducted in the statistical programming environment R (Version 4.1.0; R Core Team, 2021). Categorical datasets were simulated by transforming continuous multivariate normal variables into ordered categories. To simulate p variables in accordance to a factor model, we sampled random variables under the multivariate normal distribution $N(0_p, \Sigma)$, with 0_p denoting the vector of p variable means (all set to 0) and Σ denoting the $p \times p$ model-implied correlation matrix using the *rmvnorm* function from the R package *mvtnorm* (Version 1.1-3; Genz et al., 2021).

As for the computation of polychoric correlations, two different functions were used for binary datasets and datasets with four response categories. For binary datasets, polychoric correlations—which could also be referred to as tetrachoric correlations in the binary case—were computed with the *tetrachoric2* function from the R package *sirt* (Version 3.12-66; Robitzsch, 2022). For datasets with four categories, polychoric correlations were computed with the *polychoric* function from the R package *psych* (Version 2.2.5; Revelle, 2022). The psych implementation would have worked also for binary datasets, but the sirt implementation was preferred because it was significantly faster than the psych implementation, which greatly facilitates applications of $\text{NEST}_{\text{poly}}$ in large-scale simulations.

A common problem with the estimation of polychoric correlations are nonpositive definite correlation matrices (Garrido et al., 2013; Green et al., 2016; Timmerman & Lorenzo-Seva, 2011; Weng & Cheng, 2017). This implies that some eigenvalues of the sample correlation matrix may be negative, which contradicts their interpretation in the present context as indicators of variance explained by factors. The default setting to handle nonpositive definite correlation matrices for polychoric correlations in sirt and psych—as of their respective versions in the present work—is to apply the smoothing procedure from the psych package, which rescales estimated polychoric correlation matrices in a way that all

eigenvalues are nonnegative. We retained the default smoothing procedure in all estimations of polychoric correlations in the present work.

Erklärung an Eides Statt

Hiermit versichere ich an Eides Statt, dass ich die Dissertation mit dem Titel »Validierung und Weiterentwicklung neuer Methoden zur Bestimmung der Faktorenzahl in explorativen Faktorenanalysen« selbstständig und ohne unzulässige fremde Hilfe unter Beachtung der »Grundsätze zur Sicherung guter wissenschaftlicher Praxis an der Heinrich-Heine-Universität Düsseldorf« erstellt habe.

Ich versichere insbesondere:

(1) Ich habe keine anderen als die angegebenen Quellen und Hilfsmittel benutzt.

(2) Alle wörtlich oder dem Sinn nach aus anderen Texten entnommenen Stellen habe ich als solche kenntlich gemacht; dies gilt für gedruckte Texte ebenso wie für elektronische Ressourcen.

(3) Die Arbeit habe ich in der vorliegenden oder einer modifizierten Form noch nicht als Dissertation vorgelegt – sei es an der Heinrich-Heine-Universität Düsseldorf oder an einer anderen Universität.

Datum: 15. September 2022

Name: Nils Brandenburg

Unterschrift: